Development of small-molecule inhibitors of the initiating proteases, C1r and C1s, of the classical complement pathway

by

Denise Rohlik

December 2019

Director of Thesis: Brandon Garcia, PhD

Major Department: Biomedical Sciences

Complement is a proteolytic cascade that upon activation plays a key effector role in the innate immune system and acts to prime the adaptive immune response. During normal homeostatic events, complement is tightly regulated for its roles in immune complex clearance, lysis of target cells, opsonization, and recruitment of leukocytes and monocytes to target areas. Several endogenous regulators are responsible for the control of complement activation, however when dysregulation occurs, aberrant complement activation has been linked to autoimmune, proinflammatory, and neurodegenerative diseases, including Alzheimer's disease. Inhibition of the classical complement component C1 may ameliorate hallmarks of autoimmune and inflammatory disease.

The serine proteases within the C1 complex, C1r and C1s, are promising therapeutic targets for structure-based small-molecule drug development. We investigated the activity of a series of small-molecule compounds identified in a large-scale fragment library screen and those from a cheminformatics computational docking screen in which hit compounds were predicted to bind the C1r or C1s proteases. Using surface plasmon resonance and ELISA-based assays for hit validation, we analyzed the binding affinities and the inhibitory $IC_{50's}$ of several compounds predicted to bind and inhibit the activation of C1r or C1s in a dose-dependent manner. In this study, we have identified four lead compounds (cmp-1611, cmp-1663, cmp-1696, cmp-1827) and their 10 active structural analogues that target and inhibit C1r activation.

Given their abilities to bind and inhibit C1r and favorable physicochemical properties, our lead compounds may provide a starting point for optimizing affinity and specificity necessary for developing novel routes of therapeutic upstream complement inhibition.

Development of small-molecule inhibitors of the initiating proteases, C1r and C1s, of the classical complement pathway

A Thesis

Presented to the Faculty of the Biomedical Sciences Program

Brody School of Medicine at East Carolina University

In Partial Fulfillment of the Requirements for the Degree

Master of Science in Biomedical Sciences

by

Denise Rohlik

December 2019

Development of small-molecule inhibitors, C1r and C1s, of the initiating proteases of the classical complement pathway

by

Denise Rohlik

APPROVED BY:

DIRECTOR OF THESIS: _____
Brandon L. Garcia, PhD

COMMITTEE MEMBER: _____
Everett C. Pesci, PhD

COMMITTEE MEMBER: _____
James P. Coleman, PhD

COMMITTEE MEMBER: _____
Qun Lu, PhD

PROGRAM DIRECTOR OF
BIOMEDICAL SCIENCE PROGRAM: _____
Richard A. Franklin, PhD

DEAN OF THE GRADUATE SCHOOL: _____
Paul J. Gemperline, PhD

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

The complement system is a proteolytic cascade that upon activation plays a key effector role in the innate immune system and acts to prime the adaptive immune response. During normal homeostatic events, complement is tightly regulated for its roles in immune complex clearance, lysis of target cells, opsonization, and recruitment of leukocytes and monocytes to target areas. In healthy systems, the balance between activation of complement and its regulation maintains a fine line. Despite stringent regulation, dysregulation does occur and is implicated in a wide array of host autoimmune, inflammatory, and neurodegenerative diseases. We are working on developing small-molecule inhibitors of the initiating proteases, C1r and C1s, of the classical complement pathway to halt aberrant activation at the beginning of the cascade using both fragment-based drug discovery and cheminformatics. Our data suggest that small-molecule inhibitors prevent classical pathway activation and support the further development of C1r- and C1s- directed therapeutics in a growing field devoid of initiating protease-specific inhibitors for the treatment of classical pathway-mediated diseases.

1.1 Complement system

In the late 1800s, immune clearance of microbial pathogens was on the forefront of scientific research, and complement made its entrance as simply a "heat labile factor in blood capable of killing bacteria" that aided antibody-mediated lysis [1]. Due to its helper role in tagging antigens and microbes for lysis, it was later termed 'complement' [2]. Despite its identifying role as a key support player in immune defense, we now know that complement is far more involved in immune health than simply eliminating microbial pathogens.

In addition to "complementing" the innate and adaptive immune response, complement can coordinate a number of responses vital to maintaining host homeostasis, such as synapse maturation and plasticity, clearance of immune complexes, angiogenesis, mobilization of hematopoietic stem-progenitor cells, tissue regeneration, and lipid metabolism [3].The complement system is acutely trained to respond to non-self-antigens, cellular debris/irregular aggregates, and apoptotic cells versus healthy host cells [4].

Complement activation occurs through a variety of different mechanisms, and the activation target varies between complement components divided between one of three different complement pathways: the classical complement pathway (CP), the lectin pathway (LP), or the alternative pathway (AP). Initiating events of complement activation are pathway-specific and utilize conserved mechanisms [4].

Under normal homeostatic conditions, the alternative pathway is always active at low levels, initiated by a spontaneous hydrolytic process known as "tick-over" (Figure 1). Complement component 3 (C3), which is the central complement component of all three pathways, spontaneously cleaves its labile thioester bond, introducing a binding site for Factor B, followed by cleavage by Factor D. These cleavage events ultimately result in the generation of enzymatically active C3 convertase (C3bBb). C3 convertase cleaves C3 to C3b, which can then go on to form its own C3 convertase [5]. In doing so, complement activates in an exponential fashion by serving as its own internal amplification loop while also continuing to progress down the pathway to generate the downstream products of the alternative pathway, leading to the formation of the lytic terminal complement complex, the membrane attack complex (MAC) [4].

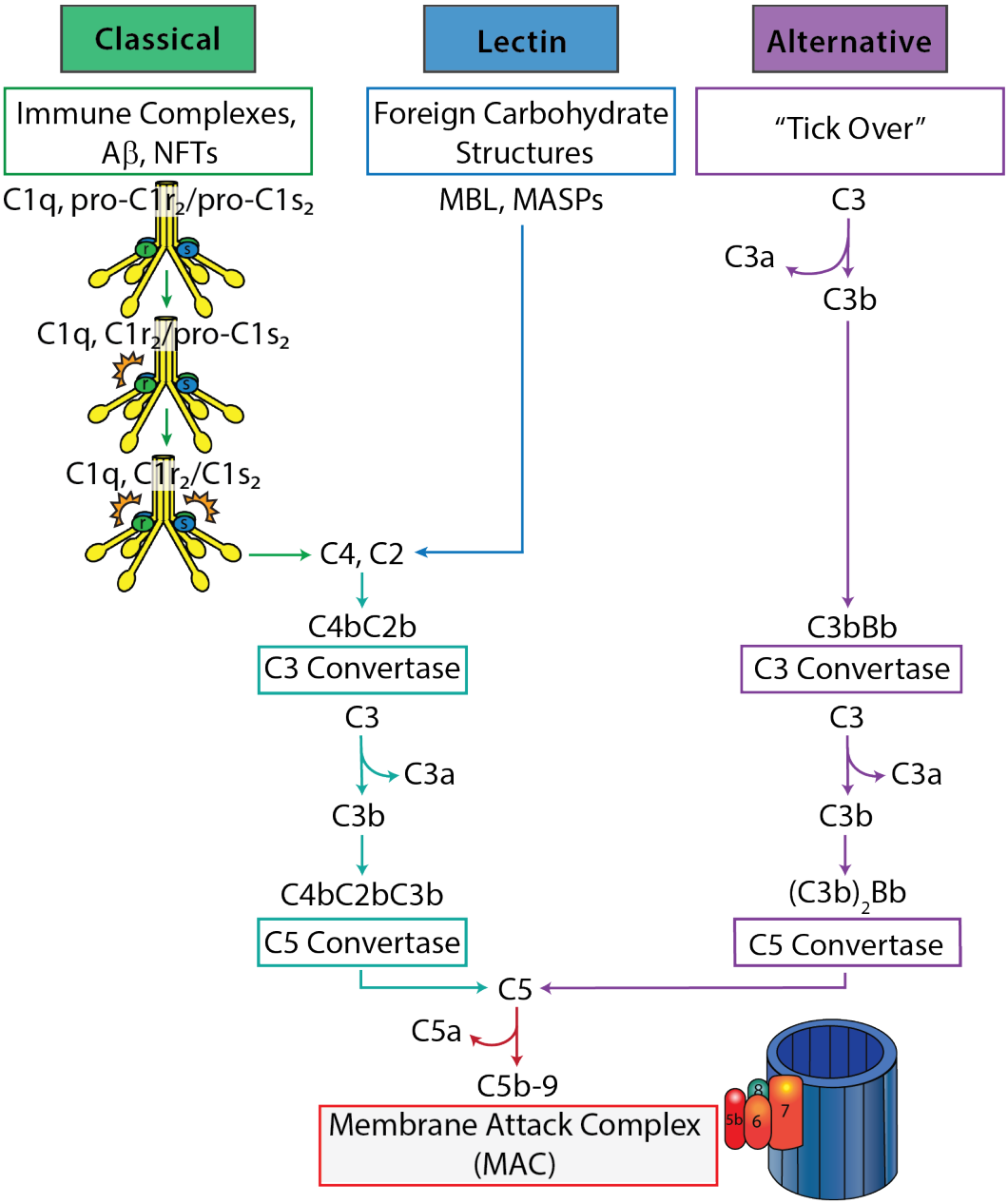

Classical
Lectin
Alternative
Immune Complexes, Aβ, NFTs
Foreign Carbohydrate Structures
"Tick Over"
C1q, pro-C1r2/pro-C1s2
MBL, MASPs
C3
C3a
C3b
C1q, C1r2/pro-C1s2
C1q, C1r2/C1s2
C4, C2
C4bC2b
C3 Convertase
C3bBb
C3 Convertase
C3
C3a
C3b
C3
C3a
C3b
C4bC2bC3b
C5 Convertase
(C3b)2Bb
C5 Convertase
C5
C5a
C5b-9
Membrane Attack Complex (MAC)
8
5b
6
7

Figure 1. The complement pathways. Complement is activated by three canonical pathways known as the classical pathway, lectin pathway, or alternative pathway. Activation of the classical complement pathway begins when the pattern recognition protein C1q binds to target surfaces resulting in the autoactivation of the zymogen C1r proteases to proteolytically cleave and activate C1s within the C1 complex (i.e., $C1qC1r_2s_2$). In a similar fashion, the lectin pathway is activated by lectin pathway-specific pattern recognition proteins in complex with mannose-binding associated serine proteases (MASPs), while the alternative pathway is constitutively activated at low levels by a spontaneous hydrolytic event known as 'tick-over.' Both the classical and lectin pathways converge at the cleavage of C2 and C4 to generate the classical/lectin pathway C3 convertases, C4b2b. Alternative pathway activation results in the formation of C3 convertases in the form of C3bBb. C3 convertases cleave the central molecule of the cascade, C3, into C3a and C3b, resulting in an amplification loop that produces increasing quantities of surface bound C3b. At high surface concentrations of C3b, C3 convertases bind an additional C3b molecule, resulting in a switch of substrate specificity to C5. Cleavage of C5 by these C5 convertases (i.e., C4b2bC3b and C3bBbC3b) results in the release of the anaphylatoxin C5a and the formation of the pore-like lytic structure called the membrane attack complex (MAC) (i.e., C5b-C9).

While the alternative pathway remains active, the lectin pathway and classical pathways are initiated by specific targets that signal danger to the host via pathway-specific recognition molecules (PRMs), mannose-binding lectin (MBL)/ficolins and C1q, respectively [4]. The recognition molecules bind the target surface, causing activation of initiating proteases, MASPs for lectin pathway or C1r and C1s for classical pathway. The two pathways progress by

proteolytic cleavage of the same downstream complement components, C4 and C2, to form the surface-bound protein complex C4bC2b. C4bC2b is the C3 convertase of the lectin pathway and classical pathways. Here, all three pathways converge at the amplification loop for the generation of C3b and C3 convertase, amplifying the outcomes of each pathway [6].

Though having similar structural appearance to C1q, MBL is activated by sugar patterns associated with bacteria, viruses, and apoptotic cells, which the lectin pathway recognizes as pathogenic [7]. C1q, on the other hand, recognizes and binds over 100 structurally diverse and distinct ligands, including IgM and hexameric IgG immune complexes [8]. Due to its diverse recognition ability, the classical complement pathway is regarded as perhaps the most complex and integrative of complement systems.

C1 complex binds a wide range of ligand activators through pattern recognition molecules, C1q (Figure 2A). The N-terminal region of C1q forms a cylinder-like collagen stem composed of interchain disulfide bridges that stabilize the structure, whereas the C-terminal region forms the six globular heads responsible for activator recognition and multivalent attachment [8]. C1q is in complex with tetrameric C1r and C1s proteases, $C1r_2s_2$. C1r and C1s are structurally homologous, being made up of the six domains: CUB1-EGF-CUB2-CCP1-CCP2-SP (Figure 2B). Each protease is characterized by the inactivated conformation of the catalytic triad, a self-inhibitory scissile loop, and cysteine salt bridge within the serine protease domain [9]. The exact orientation of trypsin-like C1r and C1s within the C1 complex is under debate, creating a rift in the theory behind the molecular events precipitating C1r autoactivation. One theory posits that the serine protease domains of each protease are condensed within the center of C1 complex, beneath the C1q stalks. The binding of C1q to a target substrate induces a conformational change in C1q collagen stems and, subsequently, a conformational change in C1r [10]. This mechanical

stress auto-activates C1r, which then acts upon C1s to induce intramolecular activation, converting C1 complex to "activated C1." The intramolecular rearrangement of the $C1r_2C1s_2$ tetramers shifts the serine protease domains outwards so that they are then accessible as binding sites [10].

The prevailing model for C1r activation places the serine protease domains of both C1r and C1s extending beyond the C1q stalks [11]. The two C1s molecules centrally dimerize via the CUB1-EGF domains, which also serve as the interface for C1r. C1q is bound to the CUB1 and CUB2 domains of C1r and the CUB1 domain of C1s. Instead of intramolecular activation, classical pathway activators concentrate C1 molecules together in a manner that aligns the exposed serine protease domains of C1r to allow for intermolecular C1 activation. If this is true, it may also be true that C1r activates neighboring C1s in the same molecule [12]. This concept is supported by biochemical evidence that lectin pathway activation occurs by proteolytic cleavage between MASPs on different pattern recognition molecules, where the role of activation-inducing glycans is to locally concentrate and orient MASPs for serine protease interaction. This process results in complement activation by the ligand-driven juxtaposition of discrete pattern recognition complexes [13].

Upon activation of C1s, activated C1 is then free to cleave target substrates, C4 and C2, to generate membrane-bound pathway amplifier, C3 convertase (C4bC2b) [14]. The classical pathway serine proteases are directly responsible for the cascade progression in the classical pathway. C3 convertases (C3bBb and C4bC2b) switch specificity to C5 upon reaching a high concentration to generate C5 convertases (C4b2b3b and C3bBbC3b). Surface-bound C5b binds C6, C7, C8β, C8αγ, and C9 consecutively forming MAC [4]. As these cleavage events occur, fluid-

phase complement protein products are also produced and act as opsonins to recruit professional phagocytes to the site for lysis of susceptible target membranes.

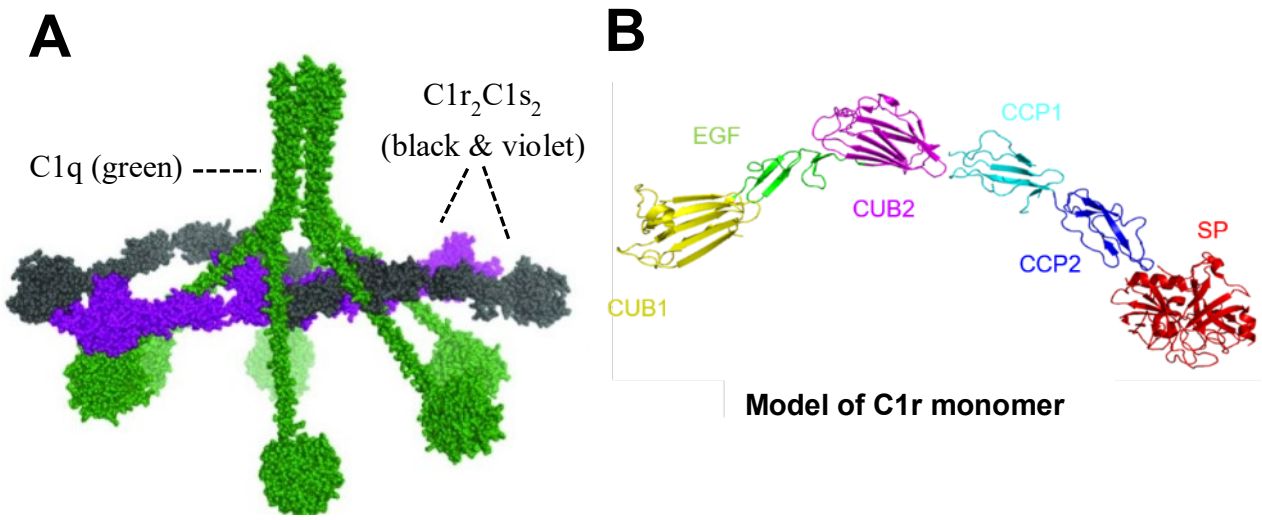


Figure 2. Structure of C1 complex and C1r monomer. (A) The C1 complex is made up of the recognition molecule, C1q, with six globular heads for multivalent attachment, and the proteolytic homodimers, $C1r_2C1s_2$. Upon C1 fixation to a target site, C1r autoactivates and cleaves pro-C1s, yielding enzymatically active C1s with its serine protease domains extending beyond the C1q stalks to provide a binding site for its natural substrates C4 and C2 [12]. **(B)** The C1r monomer is comprised of six domains, starting with the CUB1, EGF, and CUB2 at the N-terminus and the proteolytically-active CCP2 and the serine protease domains terminating at the C-terminus. The CCP2-SP domains can be recombinantly expressed to form stable and functionally active truncations [84].

1.2 Complement-mediated disorders

Complement is primarily derived from the liver to circulate systemically in plasma for continuous immune surveillance. However, there are many extrahepatic sources that produce complement components locally, such as in immune privileged sites [15]. In the brain, complement is produced by glia and/or neurons. In Alzheimer's disease, complement involved in perpetuating disease within the brain parenchyma is sourced primarily via locally produced complement [16]. Circulating plasma-derived complement is speculated to also be able to pass through the disrupted blood-brain barrier (BBB) during infection, injury, infarct or inflammation by C2a, which may enhance permeability [17].

Several endogenous regulators are responsible for the control of complement activation, but as noted by Ricklin and Lambris, "insufficient, excessive or poorly controlled complement activation can tip the balance between health and disease and lead to self-attack of host cells" [18]. Loss of function in complement components has been shown to increase susceptibility to recurrent bacterial infections, while aberrant complement activation is implicated in a host of acute and chronic diseases that aren't confined to necessarily confined to any specific tissue type or bodily region. Local and systemic activation of complement is involved in numerous autoimmune, inflammatory, and neurodegenerative diseases [3]. Due to the diverse range of ligand activators, classical complement pathway genes are shown to be upregulated in many of these disease models with frequent association with mutations in complement regulatory proteins, triggering an overactive immune cell response. Further assault and damage to host tissues by the immune system may generate even more complement-stimulating molecular patterns exacerbating the cycle of complement activation, immune stimulation, and inflammation (Figure 3) [18].

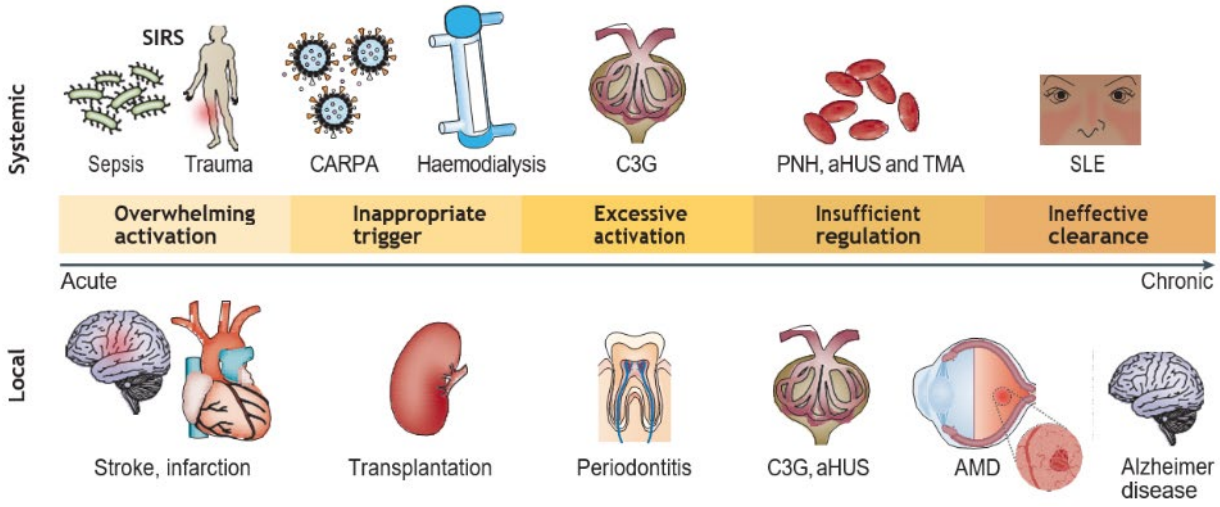


Figure 3. Involvement of classical complement activation in various disease settings. Dysregulation of the classical complement pathway can produce catastrophic disease in individual at the local or systemic level. Autoimmune, inflammatory, and neurodegenerative diseases caused by classical pathway overactivation can produce both chronic and acute disease in any bodily tissue [18]. Anti-complement drug treatment options are available for the treatment of complement 3 glomerulopathy (C3G), paroxysmal nocturnal hemoglobinuria (PNH), atypical hemolytic uremic syndrome (aHUS), and thrombic microangiopathy (TMA).

Classical pathway activation was first described for its role in antibody-mediated clearance of microbial pathogens. Though advances in high-throughput genomic analyses, new biodiagnostic algorithms, and greater understanding of structural complement mechanisms have recently unveiled that there are many other activators of the classical pathway, condensed descriptors of pathway activators is still often only associate with the classical pathway with antibody-mediated activation. Antibody-mediated immune complex formation is responsible for a broad range of diseases impacted by classical pathway amplification, though its role in each

disease setting is complex and varied. Circulating C1 binds antibody-antigen immune complexes via C1q, which is tagged by classical pathway opsonins C4b and C3b for clearance by professional phagocytes. By binding the targeted cell surface, the opsonigenic classical pathway components provide a binding site for complement receptors 1, 3, and 4 (CR1, CR3, CR4) found on erythrocytes, leukocytes, and splenic follicular dendritic cells (FDCs) for engulfment and destruction [15]. The CNS inflammatory disease, neuromyelitis optica, is initiated by IgG autoantibodies against water channel protein aquaporin-4 in complex with activated C1. Classical pathway activation drives MAC formation and lysis of oligodendrocytes and neurons [19]. In similar fashion, the autoimmune disease, bullous pemphigoid, is characterized by the activation of the classical pathway by C1q binding to IgG autoantibodies targeting the dermal-epidermal-junction, resulting in severe pruritus, urticarial plaques, and the development of tense bullae [20].

Beyond immune complex mediated activation, the classical pathway acts to prime the adaptive immune response by decorating antigens with complement fragments that induce B cell activation. Complement receptor 2 (CR2), found on B-cells, ligates to C3 degradation products (inactivated C3b (iC3b) and decaying fragments of iC3b (C3dg and C3d)) [21]. Upon protein-ligand binding, co-stimulatory signals are released that reduce the threshold for B cell receptor triggering. Follicular dendritic cells also present CR2 on their surfaces, and antigens presented by follicular dendritic cells aid in clonal selection and affinity maturation of activated B cells further amplifying B cell response [22]. Heparin-induced thrombocytopenia is a prothrombic disorder resulting from complications with heparin therapy. Platelet factor 4/heparin complexes are recognized and bound by circulating C1q, thereby activating the classical pathway. C3 degradation products bind CR2, forming platelet factor 4/heparin-B cell complexes, generating a

severe, life-threatening immune response [23]. Ischemia-reperfusion organ injuries, such as myocardial infarction, stroke, and peripheral vascular disease, are marked by tissue injury and/or death amplified by surface-bound complement fragment activity [24]. The classical pathway is thought to augment pathogenesis through C1q-IgM binding to neoantigens exposed on ischemic tissue. Studies have shown that inhibition of C1r and C1s reduce early tissue damage and recruitment/activation of dendritic cells to ischemic tissue [25].

Another complement-linked problem can occur due to complement anaphylatoxins, C3a and C5a, which stimulate T cell proliferation through T cell-expressed C5a receptor (C5aR) and C3a receptor (C3aR) signaling for the activation, differentiation, and expansion of $CD4^+$ Th1 cells, increasing the strength of T cell immune response [26]. There are numerous diseases driven by the effector role of anaphylatoxins, which is to stimulate the degranulation of endothelial cells, mast cells, and phagocytes producing an inflammatory response that can be either local, as with acute myocardial infarction [27], or systemic, such as with SIRS and sepsis [28], though the latter two involve the contribution of all three pathways.

Genetic mutations in regulatory complement proteins, such as C1INH, factor H, CD46, CD55, and/or CD59, are associated with gain of function in complement components and are causal in a multitude of complement-related disorders: C3 glomerulopathy results from complement factor H gene mutations, while paroxysmal nocturnal hemoglobinuria results from a mutation in a gene controlling CD55 and CD 59 [29,30]. Mutations in genes that encode complement regulators factor H, factor I, complement factor H-related proteins, and membrane cofactor protein and mutations encoding for complement activators complement factor B and C3 are causal in atypical hemolytic uremic syndrome [31].

Classical complement's role in neurodegenerative disease is linked to and initiated by the accumulation of cellular debris, as is seen in multiple sclerosis, schizophrenia, Parkinson's disease, and Alzheimer's disease, among others [32–35]. In the case of Alzheimer's disease, the hallmark extracellular amyloid-β peptide deposits (Aβ) and intracellular neurofibrillary tangles evoke colocalized C1q deposition independent of antibody response [36,37]. Owing to previous studies that suggest complement-mediated synaptic pruning from stages of early brain development are causally reactivated in adult Alzheimer's brains, and that, tauopathy correlates with neurodegeneration in this disease setting, it is similarly suggested that complement activation mediates neuron loss in tauopathy models [36,38,39].

Upon recognition of amyloid-β plaque and neurofibrillary tangle aggregates, C1q surface deposition occurs to initiate clearance of the peptidic debris. However, gross accumulation of cellular damage can trigger the overactivation of the classical pathway, which serves to initiate further insult to host neurons by recruitment and activation of inflammatory microglia and reactive astrocyte species (A1) to healthy tissues, especially within the hippocampus [40]. Ensuing cellular damage induces greater production of complement components by glial cells, resulting in a vicious cycle of complement activation and astrocytic degradation to host synapses [36]. In response to acute insult, astrocytes undergo transformation into a 'reactive astrocyte' state differentiated by classification as either an A1 or A2 astrocyte. A1 astrocytes are strongly neurotoxic and have been demonstrated to highly upregulate classical pathway cascade genes in conjunction with synaptic destruction, leading to the theory that complement mediates neurodegeneration in response to tauopathy and amyloidosis [40]. Indeed, when C1qa gene is selectively inactivated in transgenic murine models of Alzheimer's disease, a reduction in neuropathology and inflammation is seen in the hippocampus of late-stage Alzheimer's models

[41]. Similarly, when C3 gene is knocked out, partial rescue of synapses and amelioration of brain atrophy protects against neurodegeneration when compared to wild-type Alzheimer's disease models (Figure 4) [42]. The heavy colocalization of classical pathway components with A1 astrocytes to amyloid-β plaques and neurofibrillary tangle aggregates in the hippocampus is of great significance in dementia and Alzheimer's disease. Damage to this region detrimentally affects "memory, navigation, exploration, imagination, creativity, decision-making, character judgments, establishing and maintaining social bonds, empathy, social discourse, and language use" [43].

The classical pathway has also been implicated in Alzheimer's disease through its interactions with apolipoprotein-E (ApoE). ApoE is an endogenous protein that is primarily involved in lipid transport and cholesterol metabolization. However, ApoE plays an additional role as a potent classical pathway checkpoint inhibitor by specifically binding and inhibiting C1q with high affinity at physiological concentrations and no other complement proteins of the lectin or alternative pathways [44]. The presence of C1q-ApoE complexes observed in the human Alzheimer's disease plaques is indicative of persistent classical pathway-specific activation. Further insight into the links between ApoE and the classical pathway have revealed that Alzheimer's disease neuropathies exhibit positive correlations between lipid burden in the choroid plexus of the brain and degree of cognitive impairment [44]. These extracellular lipids are recognized by C1q in the Alzheimer's diseased brain and serve as another platform to initiate inappropriate activation of the classical pathway. High lipid burden also correlates with the ApoE4 isoform variant, which imparts a genetic predisposition to late onset Alzheimer's disease through a loss of function in the transportation and clearance of lipid deposits in the brain

yielding an appreciable connection between the ApoE genotype, classical pathway activity, and risk of developing Alzheimer's disease [44].

Given the vast nature of complement's role in multivariate disease progression, greater concerted efforts in research are being directed towards complement-targeted therapeutics. Increased understanding of disease mechanisms/triggers and complement's involvement in driving or amplifying pathogenesis underscores the importance of tailoring anti-complement therapeutics for the specific disease indication. Complement involvement in Alzheimer's disease is shown to be solely mediated by the classical pathway with early components imperative for disease progression [45]. As such, the C1 proteases, C1r and C1s, are quintessential in initiating neuropathy.

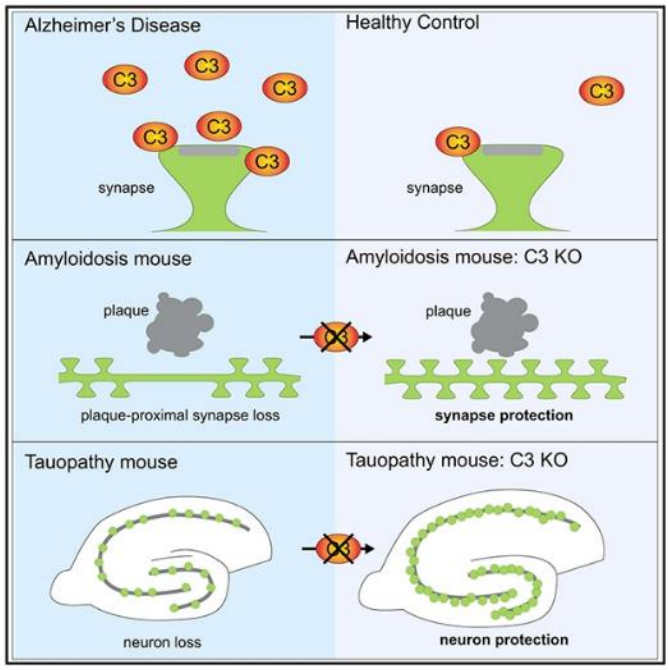


Figure 4. Complement-mediated synaptic loss in Alzheimer's models. Aberrant complement deposition products that promote pro-inflammatory immune response is hallmark of complement dysregulation in disease. In Alzheimer's disease, classical pathway components colocalize with synapses in proximity to plaques, resulting in the recruitment of neurotoxic A1 astrocytes, and resultant synaptic destruction and loss. Similarly, neurodegeneration and A1 astrocyte activity mediated by classical pathway is greatest in the presence of neurofibrillary tangles. C3 KO models demonstrate synaptic rescue and subsequent neuroprotection [36].

1.3 Complement therapeutics

The increasing awareness for the necessity of complement-based therapeutics has led to an upsurge in research and development of complement drugs. Currently, there are over 40 different published drug candidates for varying complement targets within the three pathways in

various stages of clinical development [46]. Intervention in the pathways, however, has shown to be full of complexities, and despite the significant interest, only two complement drugs have made it to market. In 2007, the FDA approved the first anti-complement drug, eculizumab (Solaris, Alexion). Eculizumab was synthesized and developed as fully humanized anti-C5 monoclonal antibody for the treatment of paroxysmal nocturnal hemoglobinuria and atypical hemolytic uremic syndrome [47]. The acquired deficiency of glycosylphosphatidylinositol in erythrocytic membranes leads to an absence of complement regulators CD55 and CD59 that prevent hemolysis by complement in individuals with paroxysmal nocturnal hemoglobinuria. By preventing the cleavage of C5, eculizumab inhibits the formation of the lytic terminal complement component, MAC, and eliminates generation of the C5a anaphylatoxin, thereby preventing lysis of erythrocytes [48]. Despite the success of eculizumab, its introduction to the market has brought to light new challenges in anticomplement drug development. One area of concern is the iatrogenic hemolysis of paroxysmal nocturnal hemoglobinuria platelets in extravascular compartments by macrophages due to C4b and C3b opsonization [30]. Additionally, during acute infections where complement activation is high, "breakthrough" hemolysis has occurred regardless of increased drug dosing . Confounding factors in disease mechanisms increase the complexity of drug development, but also highlight the need for additional disease-tailored drugs. At present, alternative anti-C5 therapeutics are a major target of pharmaceutical companies.

Inhibition of the complement pathway at the C3 level also dominates anticomplement drug research and development for its broad implication in the amplification of all three pathways, its impact on cell-mediated and humoral immunity, and the generation of anaphylatoxin, C3a, and opsonin, C3b. Currently, there are several anti-C3 drugs being evaluated

in phase II and phase III trials for treatment of issues such as ischemic disorders, tumor immunosuppression, and autoimmune disorders [49]. Yet, targeting the complement convertases located centrally and terminally in the pathways neglects diseases mediated by upstream classical pathway and lectin pathway-specific components, and due to the complexity of pathological mechanisms underlying complement-mediated disease, the need still stands for disease-specific complement regulators [50].

One drug that fills the role of targeting upstream components is C1-esterase inhibitor (C1-INH) (Cinryze). The complement drug field has also seen success with recombinant and plasma-purified C1-INH with the objective of supplementing endogenous C1-INH deficiency for the treatment of non-complement-mediated disease in patients with type I and II hereditary angioedema [51]. C1-INH is a plasma serine protease inhibitor with broad specificity beyond C1 inhibition. In addition to classical pathway inhibition, C1-INH also targets the lectin pathway, and coagulation, kinin and fibrinolytic cascades. Due to its broad specificity, C1-INH (Berinert) is also being studied for its application in other complications and disease settings where the classical and/or lectin pathway are implicated, such as delayed graft function in kidney transplantation to attenuated inflammatory markers and kidney ischemia-reperfusion injury [49]. Even with its successes, the inability to effectively act upon a restricted target presents specificity issues common across the board for the development of all serine protease inhibitors.

Serine proteases, C1r and C1s, serve as initiators of the classical complement pathway upon activation. Developing inhibitors of these proteases prevents the release of classical pathway-mediated pro-inflammatory cytokines and halts the proteolytic cascade before it can become inappropriately activated. Autoantibodies, as well as amyloid-β/neurofibrillary tangle aggregates, have been well-documented in numerous disease settings to drive classical pathway

overactivation, triggering inflammatory disease progression. Due to pathway specificity, especially in that of Alzheimer's disease, the initiating proteases serve as promising targets anticomplement targets for ameliorating pathogenesis. Currently, anti-C1q antibody, ANX005 (Annexon Biosciences), and monoclonal anti-C1s antibody, TNT009/BIV009 (True North Therapeutics), are in phase I and phase III clinical trials for treatment of Guillain Barré syndrome and cold agglutinin disease, respectively [49]. It stands to reason that their upstream classical pathway inhibition may prove to be both applicable and therapeutic in other classical pathway-mediated disease settings. However, antibody-based drugs face challenges with dosage and tissue penetrance when mediating complement in disease.

Though inhibiting the serine proteases of C1 halts only the activation of the classical pathway, leaving the other pathways open for immune protection, classical pathway complement serves an important role in elimination of bacterial pathogens. One potential problem arising from classical pathway inhibition is greater susceptibility to bacterial pathogenesis. Eculizumab trials showed that susceptibility to infection can be overcome by dosing the patient proactively and/or concurrently with constant antibiotic treatment and vaccinations against likely bacterial invaders [50]. Additionally, early classical pathway genes are also seen upregulated in early stages of Alzheimer's disease, and studies with C1q-knockout models, show marked increases in amyloidosis and tauopathy in contrast with wild-type models, implicating clearance of cellular debris as another important functional role of classical pathway [52]. However, research has shown that amyloid-β plaques and neurofibrillary tangles can persist in the brain without cognitive or motor impairment, so long as inflammation and neuronal loss is attenuated by eliminating classical pathway opsonization [36]. Complement's roles in synaptic pruning are no longer essential to adult brain development and inhibition of classical pathway conversely shows

neuroprotective activity against traumatic injury and age-related neurodegeneration [40]. Another risk that has been under discussion is that classical pathway inhibition may increase vulnerability to unforeseen other pathologies, such as systemic lupus erythematosus stemming from the genetic deficiency of C1q [53]. However, there has yet to be evidence that C1q deficiency that is acquired later in life, versus as a genetic defect, has any link to systemic lupus erythematosus or similar complications. The greatest issue that arises from targeting serine proteases is off-target risks from increased penetrance and cytotoxicity due to lack of inhibitor specificity. Currently, there are a number of drugs trying to overcome this challenge by creating small-molecule inhibitors capable of being optimized to increase specificity [54].

1.4 Small-molecule inhibitors

The traditional route for developing protease inhibitors is the identification of small-molecules that block the action of the protease by binding the active site [54]. The development of small-molecule inhibitors as a drug provides promising results for complement regulation. The low-molecular weight yields the benefit of diffusion across cellular membranes, which is especially important for transport across intestinal epithelial cells for oral drug administration, and for the potential to cross the blood brain barrier, for neurodegenerative disease targets marked by blood brain barrier injury and permeability. Larger biologics such as recombinant proteins and monoclonal antibodies are unable to freely travel across membranes and often require local injection to the disease site [50]. Additionally, Fab fragments of monoclonal antibodies have inherently short half-lives and are not a "tunable pharmacologic intervention" as are small-molecules, which also allow for better titration [55]. Small-molecules have the advantage in that they can be developed for oral bioavailability in which they are able to travel systemically. Complement has inherently high turn-over, especially of C3 due to the positive

feedback amplification loop, which allows for clearance of the small-molecule inhibitor so that complement function can be restored after the disease is cleared and the drug regime halted [50].

Complement turn-over is very rapid and during acute infection or acute disorders, complement can be produced rapidly and in great quantities in the blood. High concentrations of complement components can place a strain on dosing requirements. Typically, monoclonal antibodies must be delivered to the patient in single subcutaneous injections limited to one to two mL volume of antibody at between 100 and 150 mg/ml (approximately two - three mg/ml in humans), whereas some complement fragments, such as C5, can circulate in concentrations of 90 to 172 mg/L of blood [47]. This differential creates shortcomings in treatment of certain diseases with antibody-based drugs. Small molecules, on the other hand, can be orally dosed and concentrations titered to meet disease demands. In chronic disease models, small-molecule inhibitors may provide the best therapeutic option as an orally available drug that can be taken regularly [51].

Molecular weight of the small-molecule is an important consideration and the small size also allows for optimization of the compound to increase specificity and binding affinity to the target. Fragments, which are characterized as having a molecular weight of less than 300 Da and high solubility, have been especially efficacious in this endeavor [56]. Compounds that form aggregates in solution limit the accuracy of surface plasmon resonance (SPR) binding kinetics by causing false positives at high concentrations [57]. Utilizing x-ray crystallography to attain high-resolution crystal structures of the target protein, the molecular complexities of protein-protein interaction mechanisms, as well as critical surface target sites, can be attained to drive structure-based fragment design [58]. However, the formation of protein crystals capable of diffraction can

be challenging and required the production of pure, functional, homogeneous protein suitable for screening and structure determination [56].

Specificity has proven to be a challenge in inhibiting serine protease activity with off-target effects and toxicity being the main obstacles to overcome. Nafamostat mesilate (futhan) is a small-molecule protease inhibitor used for treatment of disseminated intravascular coagulation and acute pancreatitis that affects multiple plasma protease systems. Yet, due to its broad specificity, cytotoxicity results *in vivo*, making it a poor drug target [59]. Fragments also lack great specificity as well as low binding affinity that larger compounds may possess. However, because of these qualities, they are more likely to bind a target than a larger molecule. Moreover, fragments have the advantage of being optimized to increase those characteristics using one or more of the three existing strategies for fragment optimization – fragment linking, growing, and merging [60]. And due to their small size, researchers have more starting fragments to work with and more opportunities to make better quality decisions for generating lead compounds. Accordingly, there are many compounds currently under evaluation in clinical trials that were developed from fragments-based drug discovery [61].

1.5 Structure-based drug design

The molecular mechanisms by which complement is activated and drives pathology is critical for selecting and designing effective inhibitors of the initiating proteases. The molecular structure of enzymes is essential for deriving active sites and binding sites, and conformational changes during activation is key to design an inhibitor capable of interacting with sites that undergo those changes. Small nuances in the structure of the active site can lead to enormous implications as to what can ligate near or within the site [62]. Structure-based drug design hinges entirely on the solved, high-resolution structure of the protease, and x-ray crystallography is the

best tool to achieve that end. High-resolution x-ray crystallography defines the electron density of a given amino acid allowing for researchers to visualize bond angles/constraints and most likely conformation of rotamers, which are side-chains that adopt preferred conformations based on local energy minima [63]. Many crystal structures of various enzymes have been published along with their consequent mutations. Both C1r and C1s crystal structures have been produced and characterized for enhanced understanding of structure and activation mechanisms.

To optimize the development of our small-molecule inhibitors, we opted for structure-based drug design utilizing published crystal structures of C1r and C1s. Because we are looking to specifically target the proteolytically active serine protease domains of both C1r and C1s, we created biologically functional CCP2-SP domain truncations. Ideal crystal conditions and compound concentrations can allow for empirical binding data through effective co-crystallization of inhibitors in complex with C1r and C1s. Co-crystallization provides true binding poses and binding locations of inhibitors and opens the possibility of identifying binding sites not previously screened for [64]. Understanding these binding conformations in fragment-based drug design can also provide the basis for merging two compounds across relevant sites on the protease to increase specificity, binding affinity, and protease inhibition [65].

In one area of study, C1r inhibitors were assessed using a structure-dependent fragment-based drug discovery approach to screen compound libraries for fragments predicted to bind serine proteases based on known structures. This method allowed for identification of compound hits custom-selected for favorable physicochemical properties for serine protease inhibition. By employing this approach, hits were more frequently generated than in high-throughput screening, though they possessed relatively weak binding affinities, necessitating further optimization [55].

In a separate, orthogonal approach, we used computational cheminformatics to generate a list of priority compound leads from *in silico* libraries that index compound information to deliver predictive information about compound docking. Compounds computationally evaluated to bind within specified regions of C1s were derived from published program database (PDB) crystal structures with defined active sites. This method allowed for prioritization of hits via quantification of averaged predicted binding energies with C1s and favorable binding poses. Both methods have been used successfully in the advancement of small-molecule drug development [55,61,66].

The aim of this study was to identify small-molecule inhibitor leads of the initiating proteases, C1r and C1s, of the classical complement pathway and optimize those leads to pave the way for future drug development. We screened thousands of compounds through complex fragment-based libraries and employed cheminformatics to triage potential small-molecule leads. Using surface plasmon resonance, ELISA-based inhibition assays, enzymatic cleavage assays, hemolysis assays, and x-ray crystallography, we narrowed our leads to those with physicochemical properties that lent favorable drug-like qualities. Through our study, we have discovered four novel small-molecule compounds that display favorable binding affinities and inhibitory properties against C1r and C1s to halt the downstream progression of the classical complement pathway.

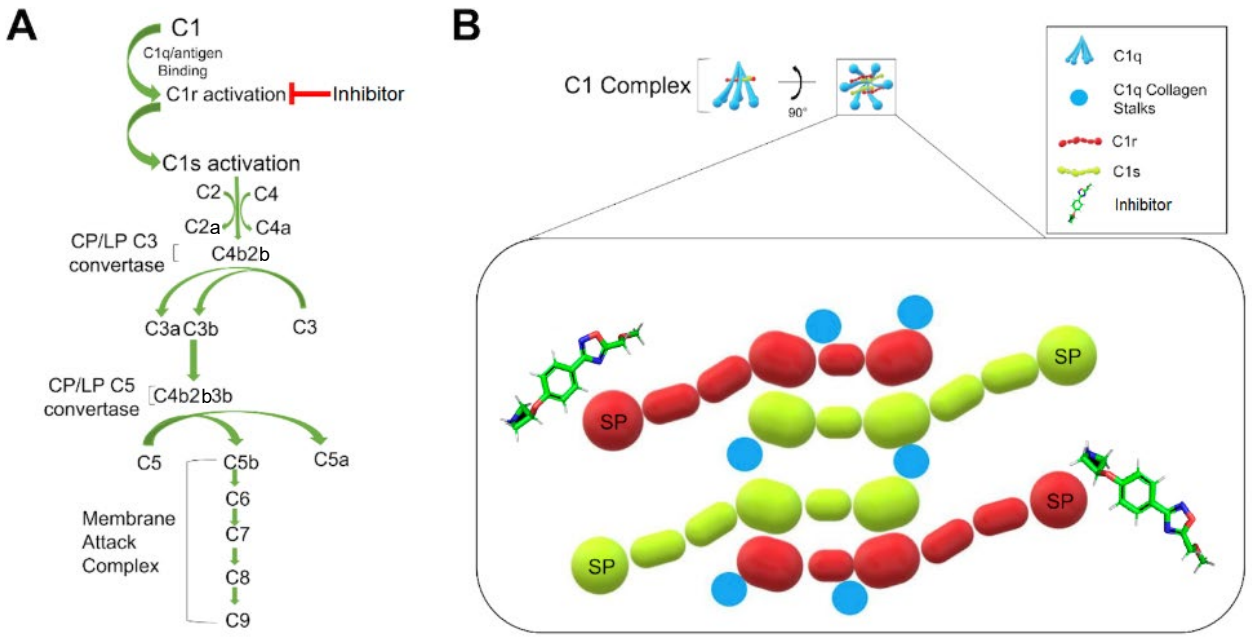


Figure 5. Pathway inhibition at the level of the initiating proteases of the classical pathway. (A) Use of a small-molecule inhibitor on C1r halts the autoactivation of C1r upon antigen recognition and binding. Downstream complement products would be significantly reduced in diseases exacerbated by classical pathway involvement and potentially arrested entirely in diseases mediated by the classical pathway. **(B)** The arrangement of C1r and C1s dimers in the C1 complex allows for ease of accessibility by inhibitors. The exposed serine protease domains on the C-terminus of $C1r_2C1s_2$ provide a favorable binding target. Figure 5 is adapted from reference [85].

CHAPTER 2: MATERIALS AND METHODS

2.1 Recombinant expression, purification, and refolding of C1r-domain truncations

Recombinant C1r-domain truncations of the CCP2-SP domains were produced by subcloning DNA oligonucleotides flanked with a 5' *Bam*HI site, a 3' *Not*I site, and a stop codon into the pT7HMT vector. The plasmids were transformed into *E. coli* BL21 and plated for incubation overnight at 37ºC. The bacterial colonies were grown in lysogeny broth (LB) (10 g/L trypton, 5 g/L yeast extract, 5 g/L sodium chloride) and transferred to Terrific Broth (TB) (11.8 g/L tryptone, 23.6 g/L yeast extract, 9.4 g/L dipotassium hydrogen phosphate, 2.2 g/L potassium dihydrogen phosphate) supplemented with kanamycin.

The cultures were induced with isopropyl β-D-thiogalactoside once they had reached an optical density of 0.6-0.8. The cultures were then incubated and stirred overnight at 37ºC and later spun at 6,000 x g for 10 minutes. The supernatants were poured off and 100 mL denaturing lysis buffer (6 M guanidine HCL, 100 mM Tris pH 8.0, 10 mM imidazole pH 8.0) was added per protein prep to the remaining bacterial pellets. The pellets were resuspended, and the lysates centrifuged at 13,000 x g for 30 minutes.

Nickel affinity columns with 10 mL nickel resin slurry were prepared by first washing with 50 mL denaturing binding buffer (20 mM sodium phosphate pH 6.0, 500 mM NaCl, 8 M urea, 10 mM imidazole pH 8.0). Lysates supernatants were dispensed onto the nickel columns and flow rate was adjusted to 1-2 drops per second. After the total lysate had passed through the column, the column was then washed with 25 mL denaturing binding buffer. Two mL denaturing elution buffer (20 mM sodium phosphate pH 6.0, 500 mM NaCl, 8 M urea, 200 mM imidazole pH 8.0) was passed through the column, and 5 mL denaturing elution buffer subsequently passed through and collected into a 50 mL conical tube at a drop rate of 1-2 drops per second.

50 mL of 2SP refold buffer (50 mM Tris pH 8.3, 3 mM reduced glutathione, 1 mM oxidized glutathione, 5 mM ethylenediaminetetraacetic acid (EDTA), 500 mM arginine) was freshly prepared and added to each 5 mL eluate sample and allowed to sit at room temperature overnight. The 50 mL 2SP samples were then transferred into 3.5K dialysis tubing and dialyzed against 4 liters of 2SP refold buffer #2 (50 mM Tris-HCl pH 7.4, 145 mM NaCl) for 4 hours. The samples were then placed in freshly remade 2SP refold buffer #2 to dialyze overnight.

The 2SP samples were concentrated down to 12 mL final volume using 3.5K MWCO centrifuge spin filters and purified by gel filtration column on FPLC. Activated (+DTT) TEV was exchanged in native binding buffer (10mM imidazole, 500 mM NaCl) and 50 μL was incubated overnight at 25^{o}C with pooled peak gel filtration fractions to remove the polyhistidine tag (His-tag).

The 2SP samples were then concentrated to 12 mL once more and run through the FPLC nickel column to collect the "untagged" protein contained in the flow-through. The pooled peak was then concentrated to a final volume yielding less than 1.5 mL of purified, refolded 2SP. To confirm proper refolding of SDS-PAGE gel electrophoresis was run with C1r-CCP2-SP under reducing and nonreducing conditions (Figure 6). Under nonreducing conditions, a band present at 38 kDa, is seen, matching the molecular weight of the C1r domain truncation, *Lane 2*. When properly purified and refolded, under reducing conditions, *Lane 4*, the disulfide bridge between α and β chains of C1r-CCP2-SP is broken resulting in the expected protein product bands at 29 kDa and 13.5 kDa.

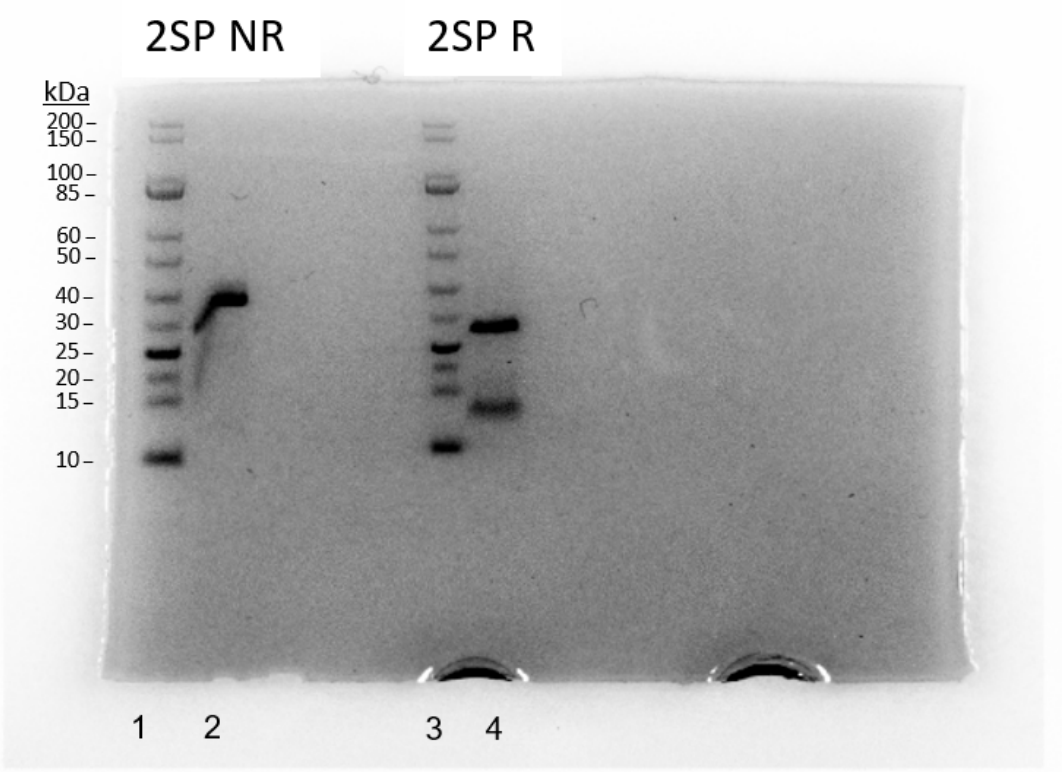


Figure 6. SDS-PAGE gel electrophoresis of refolded C1r-CCP2-SP under reducing and nonreducing conditions. The domain truncation, C1r-CCP2-SP, has a molecular weight of 38 kDa, as seen in nonreducing conditions, L*ane 2*. When purified and refolded, under reducing conditions, *Lane 4*, the disulfide bridge between α and β chains of C1r-CCP2-SP is broken resulting in the expected protein products at 29 kDa and 13.5 kDa. SDS-PAGE gel electrophoresis confirms proper protein refolding during synthesis under reducing conditions.

2.2 Compound library

The initial screen of ChemDiv compound library yielded 2,000 small-molecule or fragments hits predicted to bind C1r. Those compounds were purchased and received as powder and resuspended in dimethyl sulfoxide (DMSO) to a final concentration of 10 mM and stored at -20ºC. The structural analogues of lead compounds, cmp-1611, cmp-1663, comp-1696, and cmp-

1827 were identified in ChemDiv's compound library using structural similarity search parameters. Those analogues were also purchased, received, resuspended to 10 mM in DMSO, and stored at -20ºC.

2.3 Surface plasmon resonance

SPR experiments were conducted using a Biacore T200 (GE Healthcare) at 25ºC and HC1500M sensor chips (Xantec). All experiments were performed using a running buffer (HBS-T) of 20 mM HEPES, 140 mM NaCl, 0.005% Tween-20, 5% DMSO. At the start, finish, and every 50 cycles through the duration of the experiment, DMSO calibration curves ranging from 4 to 5.5% DMSO were collected. Over the course of this study, 13 sensorchips and 21 ligand immobilized surfaces were created in total. In all cases, C1r or C1s proteins were immobilized onto chip surfaces using amine-coupling chemistry with 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide) (EDC), N-hydroxysuccinimide (NHS), and ethanolamine. Briefly, 100 mM of EDC and NHS were mixed and injected to activate the chip surface followed by the C1r or C1s protein ligand captured at a defined resonance unit (RU) level, followed by reaction quenching with 100 mM ethanolamine pH 8.5. Flow cell 1 was always used as a reference cell (no ligand) unless otherwise noted. All sensorgrams were reference and blank subtracted and analyzed using T200 Evaluation Software (GE Healthcare).

2.4 SPR screening of C1r-binding compounds

Five μL of each of the 2,000 10 mM compound fragment hits were diluted in 95 μL of HBS-T in 96-well plates to yield 500 μM solutions with a final DMSO concentration of 5% to match the running buffer. The compounds were then visually inspected for solubility in running buffer and insoluble compounds were discarded from the "clean screen." Using an uncoupled

HC1500 SPR chip, compounds were injected over a single flow cell to assess nonspecific interactions with the chip surface. Compounds that exhibited >5.0 RU of residual binding at 10 s post injection were removed from the screen. The remaining 1,600 compounds were injected over the chip with 500 μM full-length C1r immobilized at the surface to test for binding affinity to C1r. Baseline noise was accounted for using T200 evaluation software. A theoretical maximal binding signal (R_{max}) was calculated using the equation R_{max} = (C1r immobilization level (RU) x (mol $wt_{compound}$ / mol wt_{C1r}) * n), where mol wt_{C1r} = 92,000 Da, and n is the binding stoichiometry, assumed here to be 1. Compounds were triaged as hits if all the following criteria were met: injections fell within the DMSO calibration curve, did not exhibit > 5.0 RU of residual binding to the reference surface, did not exhibit abnormal sensorgram shape, and did not exhibit superstoichiometric binding.

2.5 Evaluation of dose-dependent binding by SPR

To test for dose-dependent binding, hit compounds were injected over three replicate flow cells containing both high and low immobilization densities of C1r. Nefamostat/FUT-175/futhan (Sigma Aldrich) is a promiscuous serine protease inhibitor previously reported to bind to C1r. We obtained dissociation constants for futhan/C1r and futhan/C1r-CCP2-SP in our SPR-based assay to establish an appropriate control injection of futhan at a saturating concentration. Futhan was then subsequently used in each experiment at 10 μM to monitor inter-plate instrument and biosensor performance. Triaged compounds were tested for dose-dependent binding to full-length C1r and/or C1r-CCP2-SP using a compound concentration range of 7.8 to 500 μM. T200 evaluation software was used to calculate steady-state affinities for each compound by fitting sensorgrams from each variable concentration injection dataset using a 1:1 Langmuir model of interaction constrained by a maximal experimental binding response (R_{max})

derived from 10 μM futhan injections and corrected for the molecular weight of each individual compound.

2.6 Molecular docking

Molecular docking of compounds into the C1r domain truncation model (C1r-CCP2-SP; PDB: 1GPZ) was performed using AutoDock Vina 1.1.2. based on the given coordinates for the center of the protein structure [46, 14, 30] and size [74, 64, 54 in Å] of the target box in WinCoot 0.8.6.1. using the superpose function [67,68]. Protein structures were prepared using PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

2.7 Complement inhibition assay

High-binding 96-well ELISA plates were immobilized with 3 μg/mL IgM dissolved in coating buffer (100 mM Na_2CO_3/$NaHCO_3$, pH 9.6) dispensed in 100 μL volumes per well and incubated overnight at 25ºC. The plates were then washed three times in washing buffer (50 mM Tris pH 8.0, 200 mM NaCl, 0.05% (v/v) Tween 20) and blocked using 100 μL of 1% bovine serum albumin in PBS-T for one hour at 37ºC. The plates were washed three times with washing buffer, and in a non-binding 96-well plate, two-fold dilutions of our 10 mM compounds were prepared using 5 μM compound in 95 μL classical complement ELISA buffer (20 mM HEPES (pH 7.3), 0.1% (w/v) gelatin, 140 mM NaCl, 2 mM $CaCl_2$, 0.5 mM $MgCl_2$) and 1.05% normal human serum. Row G was utilized as a positive control for full classical complement activity, and Row H was utilized as a negative control for no classical complement activity. The samples were transferred to the blocked ELISA plates and incubated for 1 hour at 37ºC. The plates were then washed three times in wash buffer and 100 μL anti-C4 antibody diluted 1:300 in PBS-T was dispensed into each well. The plates were incubated for 1 hour at 37ºC, washed three times in

wash buffer, and filled with 100 µL of HRP-conjugated goat-α-mouse secondary antibody diluted 1:3000. The plates were rocked for 1 hour at 25ºC, washed three times, and then filled with 100 µL 1-step Ultra TMB. The plates were rocked for 15 minutes in the dark at 25ºC. The reaction was halted with 50 µL of 0.16 M sulfuric acid and the plate read at 450 nm using an EnSight multimode plate reader. Positive controls (1% serum, no inhibitor) were defined as 100% C4b signal whereas negative controls (0% serum) were defined as 0% C4b signal for each column.

Compounds that inhibited C4b deposition relative to the control were then evaluated using the same assay setup across a two-fold variable concentration curve of each compound (7.8 to 500 µM). These data were used to obtain half-maximal inhibitory concentrations (IC_{50}'s) by fitting dose-response curves using inhibitor vs. response models in GraphPad Prism 7. All experiments were performed in triplicate.

2.8 C1r and C1s enzyme assay

Individual stock solutions of 1 mM of Z-Gly-Arg-sBzl (MP Biomedicals) and 5,5′-Dithiobis (2-nitrobenzoic acid) (DTNB) (Sigma-Aldrich) were made by dissolving powder in DMSO to give 10 mM stocks of each compound which were then diluted to 1 mM in assay buffer (50 mM tris-HCl pH 7.5). Full length C1r or C1r-CCP2-SP were dissolved in assay buffer to give stocks of 20 nM and 15 nM respectively. The 100 mM compound stocks (in DMSO) were pipetted in triplicate in 2.5 µL volume into a 96-well plate. Full-length 20 nM C1r or 15 nM C1r-CCP2-SP solutions were added to the wells in 25 µL quantities, along with 12.5 µL assay buffer. In a separate reservoir, equal volumes of 1 mM DTNB and 1 mM Z-Gly-Arg-sBZL were mixed, and 10 µL was added to each well to initiate the reaction. Plates were read at 450

nm using an EnSight multimode plate reader (Perkin-Elmer) every 30 seconds for 30 minutes after the reaction was initiated. Positive controls (no inhibitor, enzyme only) and negative controls (no enzyme) were defined as 100% and 0% signal for each column, respectively. For C1s, the assay buffer was HBS-T (20 mM HEPES, 140 mM NaCl, 0.005% Tween-20) plus 5 mM $CaCl_2$. A final C1s concentration of 5 nM was used for the reaction and Z-L-Lys-sBzl (MP Biomedicals) was used as a substrate at a final concentration of 100 μM.

2.9 Compound similarity analysis

Cheminformatics was employed independently of the compound library screen using Chemalytic's *in silico* approach to identify highly-specific compounds predicted to dock near or within the C1r-2SP catalytic cleft through the published C1r-domain truncation crystal structure. Compounds are identified from the ZINC database and docked in favorable conformations and their theoretical binding affinities calculated and prioritized.

2.10 X-ray crystallography

Using a 3K MWCO spin filter, C1r2SP was spun at 13,000 x g for 5 minutes to concentrate the sample volume down to 100 – 200 μL and a final concentration of at least 3 mg/mL, after correcting for the extinction coefficient. The protein solution was pipetted in a crystal sample tube and with the lead compound and dispensed onto the tray at a 1:1 mix of protein to mother liquor in the top well and a mix of 2:1 in the bottom well. Several crystallization conditions were identified with two conditions producing the most promising crystals to work with. The first condition contained 0.2 M sodium acetate trihydrate, 0.1 M Tris hydrochloride pH 8.5, 30% w/v polyethylene glycol 4,000 and the second contained 0.2 M magnesium chloride hexahydrate, 0.1 M Tris hydrochloride pH 8.5, 30% w/v polyethylene glycol 4,000. Small plate clusters reproducibly appeared within 24 hours and were harvested and

cryoprotected with supplementation of 5% glycerol to the precipitant solution. Crystals in this condition grew in the space group P $2\ 2_1\ 2_1$ with one copy of the C1r-CCP2-SP molecule in the asymmetric unit diffracting at 1.8 Å.

2.11 C1s in silico docking experiment

Using published crystal structures of C1s, Chemalytic's software utilized AutoDock Vina to computationally screen 4.27 million compounds from the ZINC database for predictive binding to the C1s dimer interface, C1s/C1q interface, C1s/C1r interface, and the C1s scissile loop. The compounds were overlaid within each of the four sites and their calculated C1s-binding energies were ranked. Compounds that scored 4.5 standard deviations above the average docking score underwent further physicochemical analysis to calculate the xLogP values of each compound and ranked accordingly. Compounds with favorable theoretical C1s-binding affinity, physicochemical properties, and diverse chemical structures were triaged.

2.12 SPR screening of C1s-binding compounds

Lead compounds predicted to bind with high affinity near or within the C1s catalytic cleft were commercially synthesized. The compounds were then resuspended in 100% DMSO to yield a final concentration of 10 mM. The resuspended compounds were further diluted in HBS-T in 96-well plates to yield 500 μM solutions with a final DMSO concentration of 5%. The compounds were then injected over the sensor chip with a concentration range of 7.8 to 500 μM and assessed as to whether they bound 500 μM immobilized full-length C1s and/or C1s-CCP2-SP. T200 evaluation software was used to calculate steady-state affinities for each compound by fitting sensorgrams from each variable concentration injection dataset using a 1:1 Langmuir model of interaction constrained by a maximal experimental binding response (R_{max}) derived

from 10 μM futhan injections and corrected for the molecular weight of each individual compound.

2.13 Complement inhibition assay

C1s inhibition was measured using the identical ELISA protocol used for testing C1r inhibition.

CHAPTER 3: FRAGMENT-BASED DRUG DISCOVERY APPROACH

3.1 Small-molecule library design

The small-molecule library design originated from a compound library of a custom-selected, commercially synthesized, 2,000 small-molecule library (ChemDiv Inc.), and was chosen to encompass five subcategories based on chemical scaffolds, functional groups, substituent moieties, stereochemistry, conformers, and chains (Figure 7). The categories were defined as: i) 250 "two-dimensional fragments" (2D-FL), ii) 250 "three-dimensional fragments" (3D-FL), iii) 250 "serine-protease inhibitor" compounds (SPI), iv) 250 "natural product scaffolds" (NPB), and v) 1,000 "protein-protein inhibitor" compounds (PPI).

Compounds in the 2D-FL library were small-molecule fragments that ranged in molecular weight between 99-330 Da, whereas those found in the 3D-FL category were fragment-sized compounds ranging in size from 109-372 Da with increased three-dimensional character as judged by increased sp^3 hybridized carbons, which lends three-dimensionality and complexity to the library (44). Small molecules from the SPI category were selected based on scaffold similarity to known serine protease inhibitors and ranged in size from 164-575 Da. Compounds from the NPB category ranged in molecular weight from 221-568 Da and were chosen for their scaffold structural similarity to natural product-derived compounds. The PPI library was made up of compounds with physicochemical properties similar to those of known protein-protein inhibitors. The compound categories were selected for their diverse and potential inhibitory qualities of C1r for future studies.

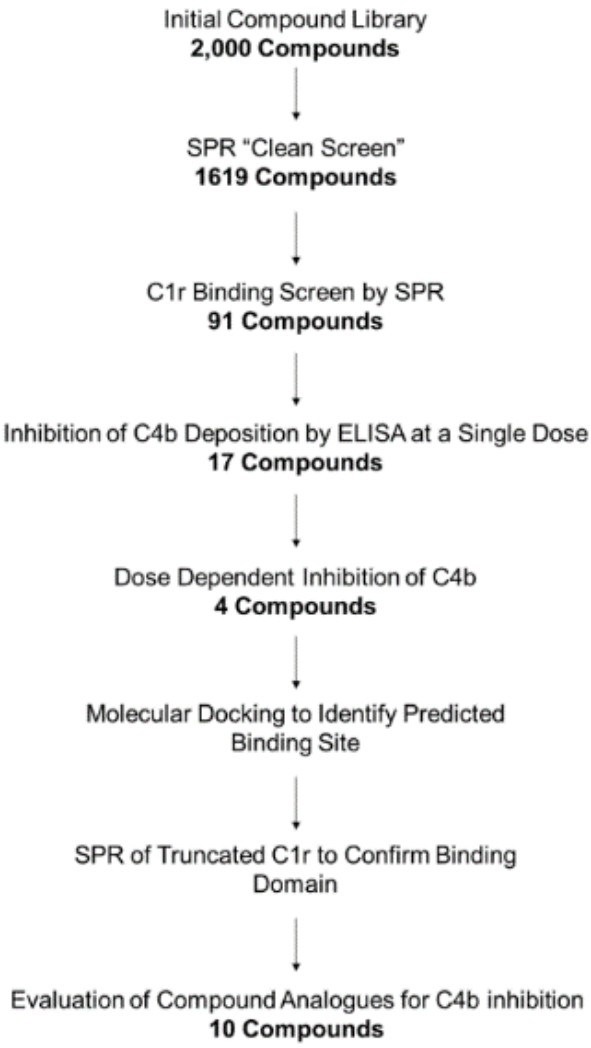


Figure 7. Study workflow of fragment-based drug discovery approach. Identification of small-molecule leads targeting C1r-CCP2-SP domain originated from a screen of commercially available fragments. Using stepwise biochemical and functional validation assays, the list of compound leads was narrowed from 2,000 fragments to four that displayed dose-dependent binding and inhibition. By identifying key structural components of each compound, an additional 10 analogues were found to exhibit similar functional activity in classical pathway inhibition.

3.2 Initial library screening of C1r-binding by SPR

Compounds that were selected for in the initial compound library screen were further triaged in an SPR "clean screen" to eliminate those with non-specific binding characteristics and/or poor solubility in aqueous SPR buffers (Figure 4). To first determine solubility, compounds were diluted to a final concentration of 500 µM in SPR running buffer (20 mM HEPES (7.3), 140 mM NaCl, 0.005% Tween-20, 5% DMSO). Compounds that were visibly insoluble at 500 µM were eliminated from future testing. Compounds that were soluble underwent additional screening for non-specific binding behavior by injecting the 500 µM compound over a blank flow cell on the SPR sensorchip for 30 seconds. Those that displayed a residual binding signal of > 5.0 RU at 10 seconds post injection were eliminated from future testing due to non-specific binding.

Of the 2,000 compounds that were screened, 1,619 passed the clean screen for further testing (Figure 8). Full-length C1r was covalently immobilized to the SPR sensorchip surface for the C1r binding screen by SPR and the remaining compounds were injected over the surface at 500 µM. Binding curves of compounds that were outside of the DMSO solvent correction curve and those that exhibited superstoichiometric binding were eliminated from future testing. Compounds selection was further narrowed by theoretical maximal binding response. Those that showed maximal binding response greater than 60% of the theoretical maximum were prioritized. From the 2,000 compounds that were originally screened, 91 compounds (4.5%) were selected for future testing based on solubility and C1r-specific binding response.

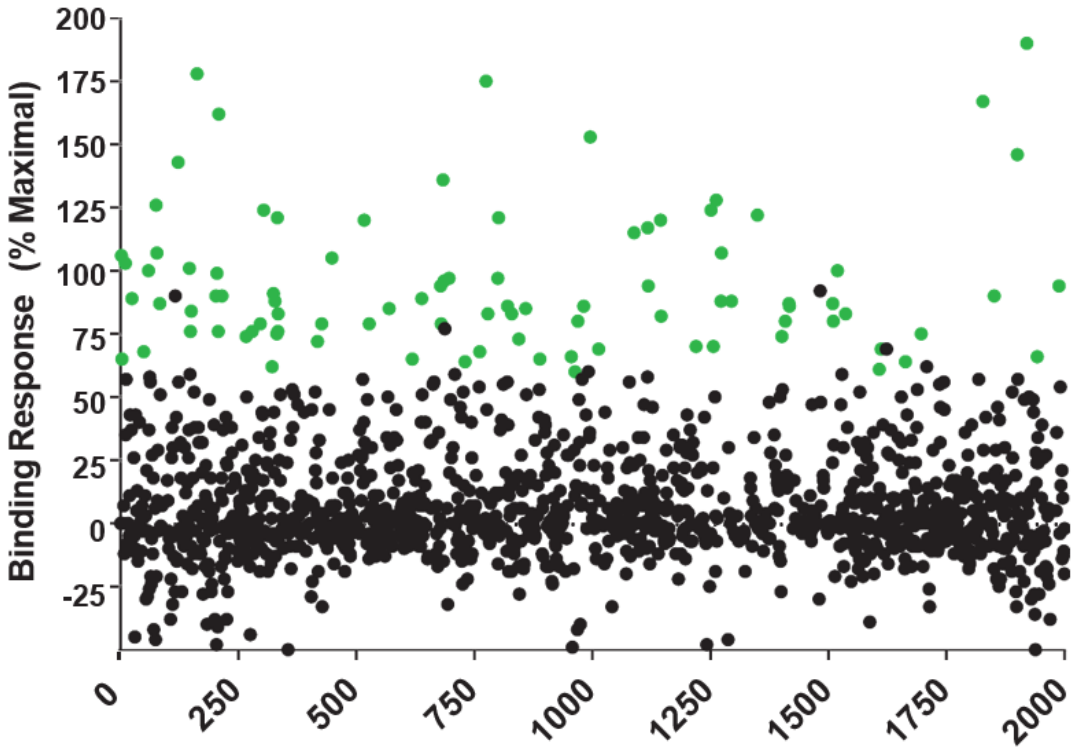


Figure 8. SPR binding response analysis of all compounds injected over full-length C1r. Compounds identified from the fragment screen were evaluated for aqueous solubility and non-specific binding via injection over a blank SPR sensorchip. Those that displayed poor solubility and/or non-specific binding were eliminated from the study. Remaining compounds were injected over a sensorchip at a concentration of 500 μM with immobilized full-length C1r. Using the calculated theoretical maximal binding response (R_{max}), compounds that presented superstoichiometric binding (i.e. >2x R_{max}) were eliminated, while those that achieved ≥ 60% R_{max} and were retained (green).

3.3 C1r-binding and classical pathway inhibition properties of prioritized compounds

The prioritized compounds were further assessed for their abilities to both bind and inhibit C1r autoactivation (Figure 9). The binding affinities of 59 of the 91 compounds were then examined for dose-dependent binding characteristics in subsequent SPR analysis by injecting two-fold variable concentration series of each compound over immobilized C1r, and they ranged in concentration from 17 to 7,480 μM. The steady-state affinities were calculated from three independent injection series.

An ELISA-based *in vitro* assay was used to determine classical pathway inhibition at a single dose by measuring downstream complement product, C4b, well deposition. Compounds were utilized in 500 μM final concentrations and optical densities (ODs) were obtained, quantifying the extent of classical pathway inhibition. Seventeen compounds displayed statistically significant complement inhibition and when serially diluted from 500 μM to 7.82 μM final concentration, 15 of those compounds were also shown to display dose-dependent binding of C1r by SPR. Those compounds were selected to undergo further testing.

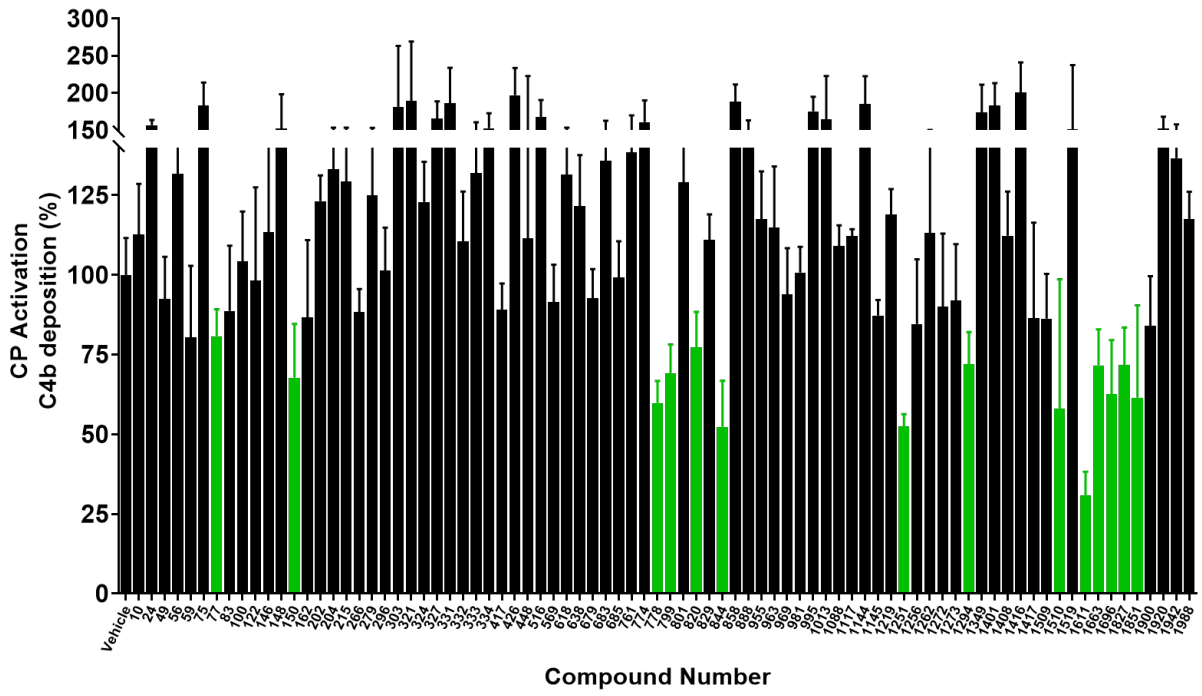


Figure 9. ELISA-based assay for classical pathway inhibition by compounds. Compounds that bound in the SPR evaluation screen were assessed in triplicate at a concentration of 500 μM for classical pathway inhibition by measuring C4b deposition in ELISA-based assay. In total, 17 compounds were identified as classical pathway inhibitors, relative to the vehicle control (DMSO), as seen in green. C4b deposition reduction was statistically significant in inhibitory compounds using the Student's t-test ($p < 0.05$).

3.4 Identification of structurally distinct C1r-binding and inhibitory compounds

Using the ELISA-based *in vitro* assay, the remaining 15 compounds were prioritized based on dose-dependent inhibition of the classical complement pathway. Of the 15 compounds, four yielded reproducible lead dose-dependent C1r-binding and inhibition (Figure 10A-D). These lead compounds, cmp-1611, cmp-1663, cmp-1696, and cmp-1827, were all derived from

the protein-protein inhibitor library, yet they remain structurally distinct from one another as small-molecule fragments. Compounds cmp-1611, cmp-1696, and cmp-1827 each exhibit favorable "rule-of-three" fragment physicochemical properties (i.e., ≤ 300 Da, ≤ 3 hydrogen bond donors, and ≤ 3 cLogP) (45). The remaining lead compound, cmp-1663, is larger than a typical fragment at 376 Da, but exhibits drug-like "rule-of-five" properties (i.e., ≤ 500 Da ≤ 5 hydrogen bond donors, ≤ 10 hydrogen bond acceptors ≤ 5 cLogP) (46). The compounds displayed C1r-binding affinities (K_D) ranging from 115 to 202 μM and C1r-inhibition values, represented by IC_{50} values, of 550 to 1,060 μM. Due to their favorable properties, these compounds were obtained in greater quantities for further evaluation.

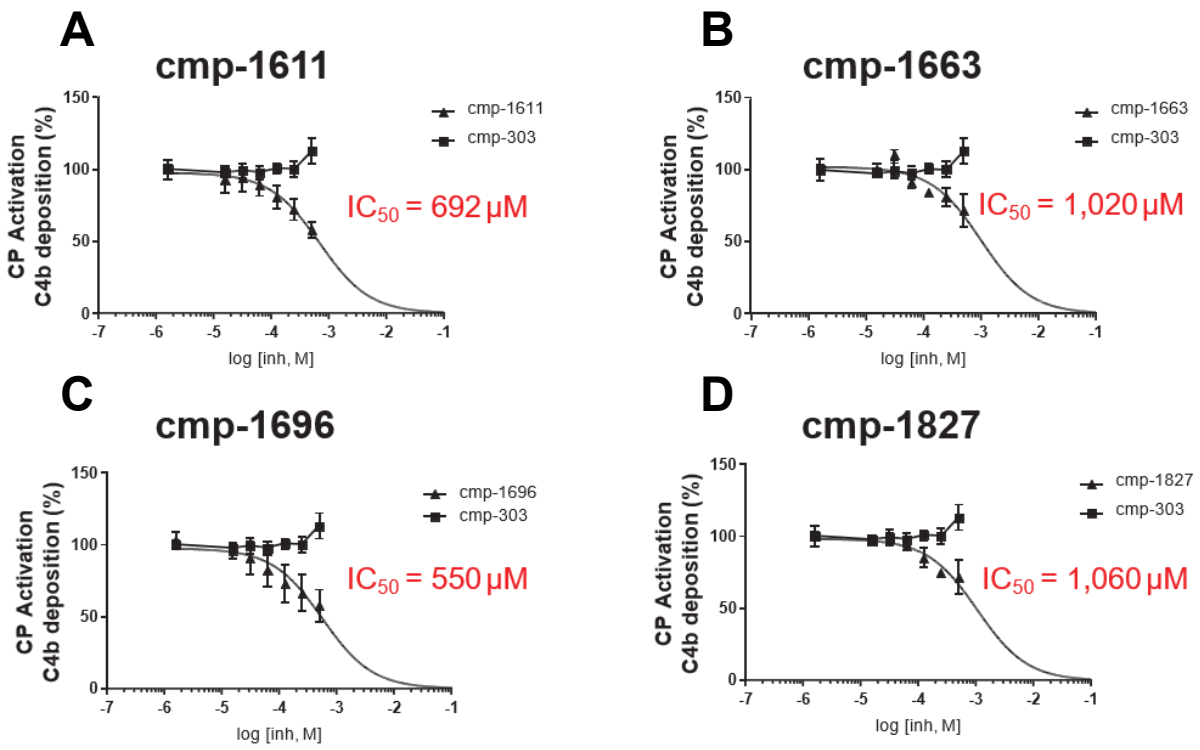


Figure 10. ELISA-based assay for dose-dependent inhibition of CP. (A-D) The 17 compounds were evaluated for dose-dependent inhibition of the classical pathway by utilizing two-fold serial dilutions of inhibitor with concentrations ranging from 10 mM to 312.5 µM. C4b deposition was used as a measure of classical pathway activity. In total, four compounds, cmp-1611, cmp-1663, cmp-1696, and cmp-1827 dose-dependently inhibited the classical pathway. Using non-linear regression analysis, inhibition curves were generated and IC_{50}'s were calculated using a 95% confidence interval. Cmp-303 is plotted as a basis of comparison, as cmp-303 was identified as a C1r dose-dependent binding compound that displayed no classical pathway inhibition ($K_{D\text{-}C1r/cmp\text{-}303}$ = 150 µM).

3.5 Mapping the binding site of each lead fragment on C1r

Following confirmation of C1r binding and inhibition, we sought to identify the binding site of each lead compound on C1r. In order to determine the binding site, known C1r-domain truncations, based on solved atomic resolution crystal structures, were utilized in three independent *in silico* docking experiments as the full-length crystal structure of C1r has not yet been solved (Figure 11A). To isolate both binding site and C1r domain, the monomeric structure of the N-terminal region of C1r (CUB1-EGF-CUB2; PDB: 6F39) [11] or the C-terminal region of C1r (CCP1-CCP2-SP; PDB: 1GPZ) [69] and CCP2-SP (PDB: 1MD8) [70] was applied to AutoDockTools1.5.6 and each lead compound docked (Figure 11B).

The predicted binding site within the N-terminal region remained within the same pocket on the CUB1 domain, whereas the binding site on the C-terminal region was within the proteolytically active serine protease domain, near or within the catalytic cleft or substrate specificity pocket S1 site. The binding affinities within each domain were predicted to be within ≤ 0.3 kcal/mol for each compound and each corresponding binding site.

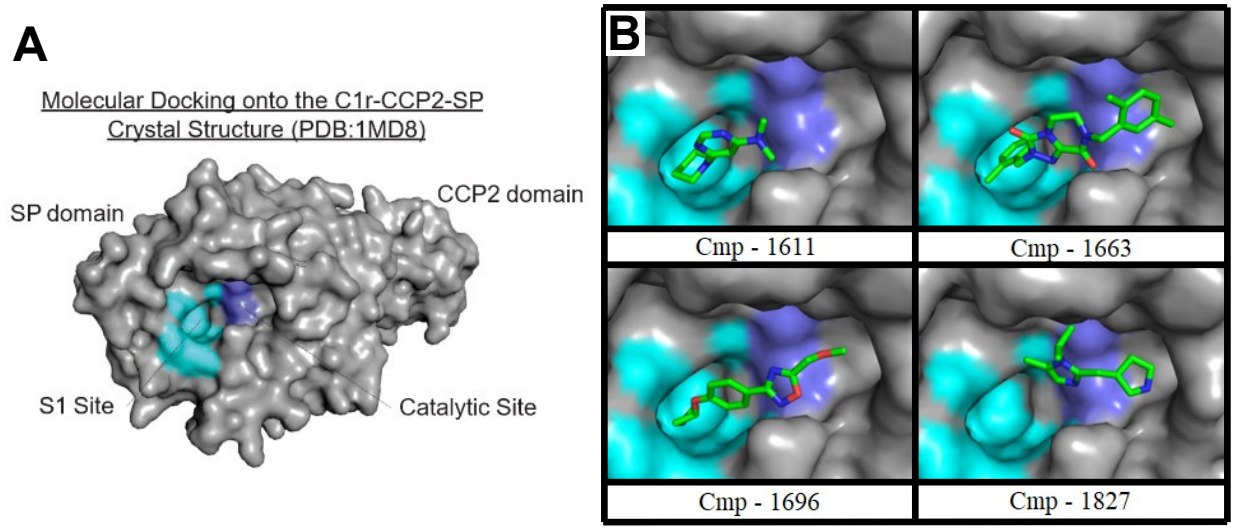


Figure 11. Crystal structure of C1r-CCP2-SP provides *in silico* predictive molecular docking data for lead compounds. (A) Using the published crystal structure of C1r-CCP2-SP domain truncation mutant (PDB: 1MD8) and known sequences of serine proteases and C1s active sites, we were able to establish the substrate specificity pocket S1 site (cyan) and catalytic triad (slate blue) of the C1r serine protease. **(B)** With the published structure of C1r-CCP2-SP, compounds were docked according to the lowest energy docking pose. All four compounds docked into the S1 site, the catalytic site, or across both sites.

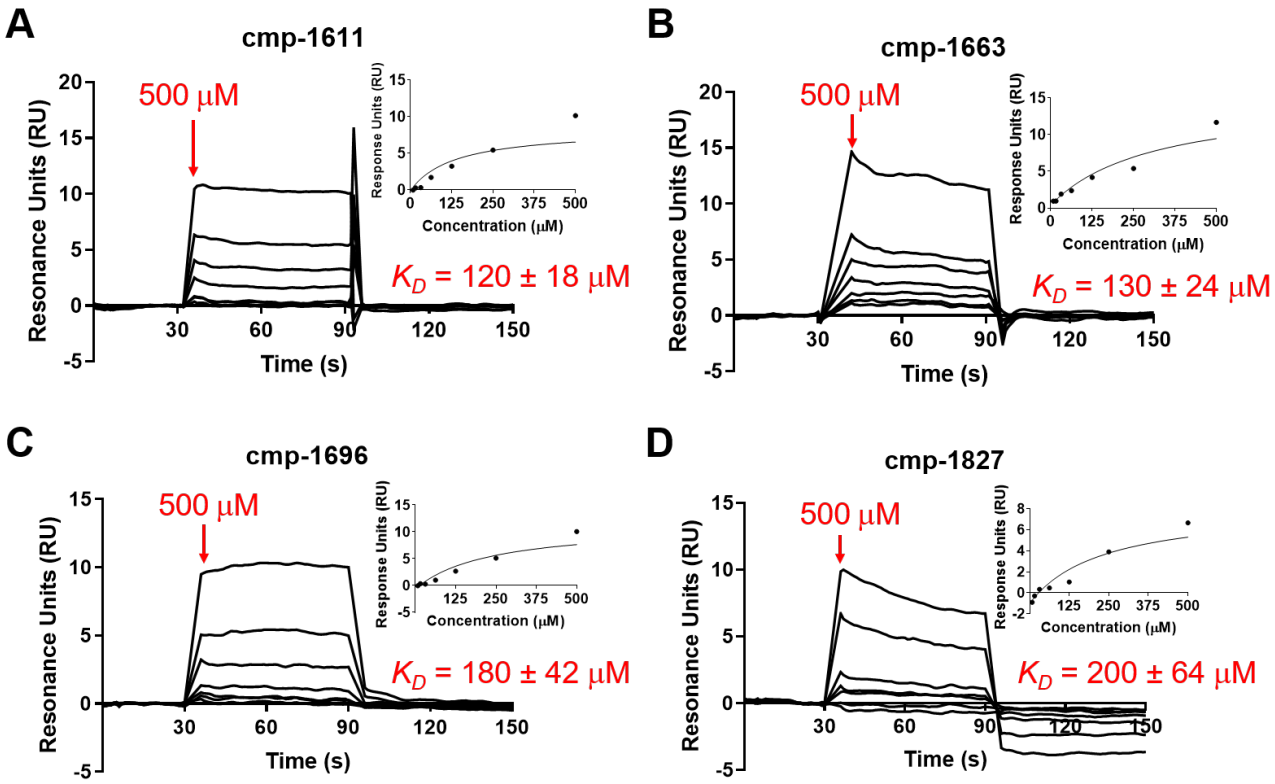
To isolate the competitive and preferred binding site of each compound, we produced the recombinantly expressed C1r-domain truncations C1r-CUB1 and C1r-CCP2-SP (we were unable to solely produce the C1r-SP domain). Those C1r-domain truncations were then immobilized separately on SPR sensorchips and the lead compounds individually injected over each surface.

The SPR binding experiment yielded differing binding affinities between C1r-CUB1 and C1r-CCP2-SP than that of what was predicted in the *in silico* docking experiments, with the lead

compounds strongly favoring C1r-CCP2-SP over C1r-CUB1. Compounds cmp-1663 ($K_{D,\,C1r}$ = 127 μM vs. $K_{D,\,C1r\text{-}CCP2\text{-}SP}$ = 137 μM), cmp-1696 ($K_{D,\,C1r}$ = 183 μM vs. $K_{D,\,C1r\text{-}CCP2\text{-}SP}$ = 211 μM), and cmp-1827 ($K_{D,\,C1r}$ = 202 μM vs. $K_{D,\,C1r\text{-}CCP\text{-}SP}$ = 213 μM) all bound with similar affinity to both C1r and C1r-CCP2-SP, while cmp-1611 ($K_{D,C1r}$ = 115μM vs. $K_{D,C1r\text{-}CCP\text{-}SP}$ = 694 μM) bound C1r-CCP2-SP with ~6-fold weaker affinity as compared to full-length C1r (Figure 12 A-H). On the other hand, cmp-1611, cmp-1663, cmp-1696, and cmp-1827 bound C1r-CUB1 with 28-fold, 176-fold, 29-fold, and 9-fold lower affinity compared to full-length C1r, respectively.

The SPR binding experiments provided evidence that the catalytic cleft, substrate specificity pocket S1 site, or both sites of C1r-2SP, is the preferred binding site of the four lead compounds. Due to the binding proximity of these compounds to the proteolytically active site of C1r, we became interested in determining whether binding directly blocked C1r activation.

SPR: Full-length C1r dose-response



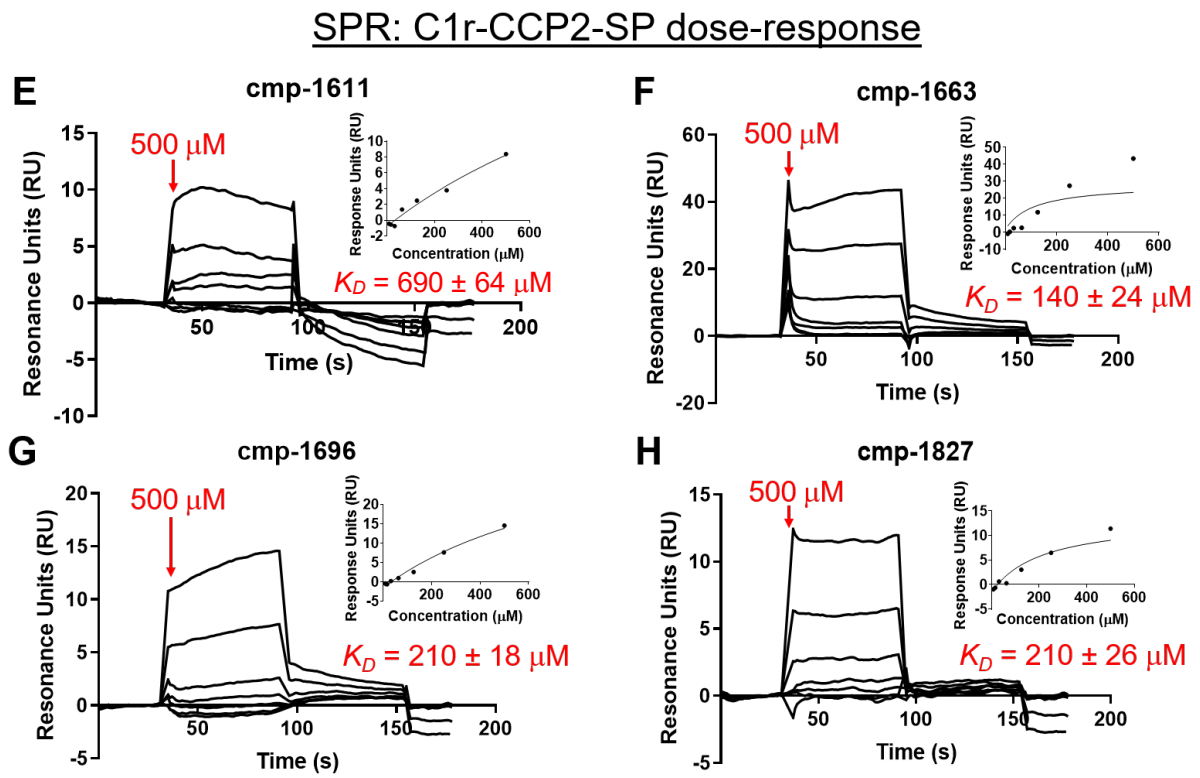


Figure 12. Domain mapping of dose-dependent binding of four lead compounds by SPR. **(A-C)** The four lead compounds were injected at 500 µM final compound concentration over immobilized full-length C1r. Dose-dependent binding response was measured using variable concentrations. Steady-state affinities and K_D's were calculated from the resulting data. **(E-H)** The ability of the compounds to dose-dependently bind C1r-CCP2-SP domain truncation mutant was assessed using identical means as full-length C1r.

3.6 Enzymatic cleavage inhibition assay

To determine whether C1r activation is directly blocked by our lead compounds, we utilized pre-activated full-length C1r enzyme, C1r-CCP-2SP enzyme, or subsequent classical complement pathway protease C1s enzyme, in an enzyme inhibition assay that is measured by

colorimetric change of *Ellman's* reagent (5,5'-*dithiobis*-(2-nitrobenzoic acid) or DTNB) marked by enzymatic cleavage of synthetic peptide substrate Z-Gly-Arg-SBzl. The compounds were utilized at a final concentration of 5 mM, (except for cmp-1663, which is insoluble at this concentration in reaction buffer and thus excluded) though, only cmp-1696 exhibited inhibition of C1r and C1r-CCP-2SP enzymes but not C1s.

3.6 Activity of structural analogs

Having previously identified molecularly distinct core structures, we sought to identify structural analogues of our lead compounds to find other dose-dependent binding and inhibitory compounds of C1r (Figure 13). Thirty-two additional compounds were generated from a structure similarity search and obtained for further testing with four analogs deriving from cmp-1611, ten from cmp-1663, 14 from cmp-1696, and four from cmp-1827. Utilizing previously employed SPR binding experiments and ELISA-based assays, six of the 32 exhibited similar IC_{50}'s and K_D's to the original compounds, marking the identification of six new, novel compounds (Figure 14). Two derived from cmp-1611, one from cmp-1663, one from cmp-1696, and two from cmp-1827.

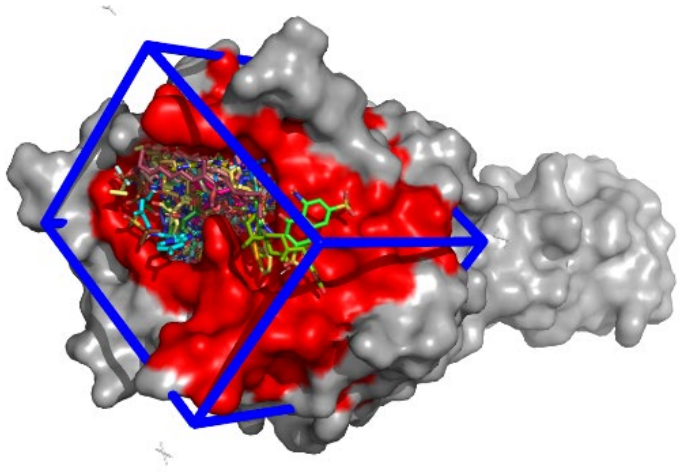


Figure 13. Predicted binding site of compounds and analogues within or near the catalytic cleft of C1r. The published crystal structure of the serine protease domain of C1r was utilized to generate the most probable docking poses of the active analogues of our four lead compounds within the space defined by the highlighted surface of interest (red) within the boxed region of the serine protease domain. The substrate specificity S1 pocket and catalytic triad are present within the specified area, and all compounds and their respective analogues were predicted to bind within this area.

Further comparison of their core structures with those of their corresponding lead compounds, yielded the identification of structurally diverse common core fragments belonging to each group (Figure 14). In the case of cmp-1611, each active compound contains a central 4,6-dimethylpyrimidine core fragment, while the two active cmp-1663 compounds are isomers differing in the position of a methyl group on a benzene ring. The active cmp-1696 analog (cmp-1696-7) is a derivative of cmp-1696 with an additional ethylpyridine substituent, while the active cmp-1827 compounds all contain a central 4-methylimidazole ring.

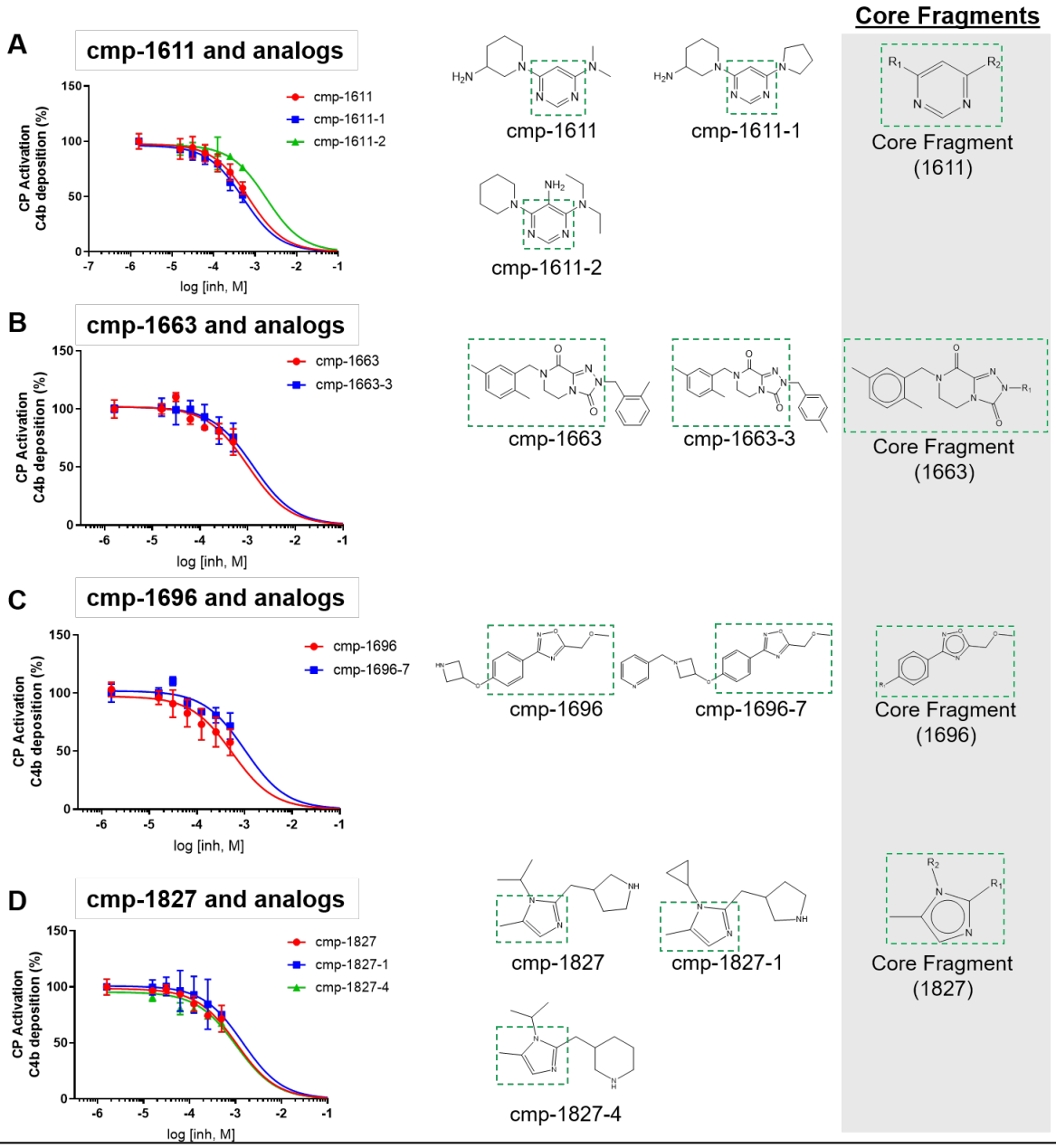


Figure 14. Dose-dependent inhibition of lead compounds and their analogues. (A-C) Structural analogues of our lead compounds were determined using a structure similarity search of cmp-1611, cmp-1663, cmp-1696, and cmp- 1827. *In vitro* ELISA assays measured C4b deposition as a basis for classical complement inhibition for compounds and their analogues. Two cmp-1611 analogs, one cmp-1663 analogs, one cmp-1696 analog and two cmp-1827 analogs exhibited dose-dependent binding and complement inhibition.

3.7 Activity of in silico docking hit fragment compounds

In a separate and independent *in silico* docking experiment, Chemalytics was employed to identify compounds *de novo* from the ZINC database predicted to dock near or within the C1r-CCP2-SP catalytic cleft through the published C1r-domain truncation crystal structure. Compounds were ranked based on their theoretical binding affinities and favorable binding poses, and from that experiment, 756 unique compounds were identified. Eighty showed favorable physicochemical binding properties, and twenty-one of those compounds were identified to share the same core fragment structures as our four compounds (Figure 15 A-B). This discovery provided further evidence for the binding and inhibitory qualities of the core scaffolds of our lead compounds.

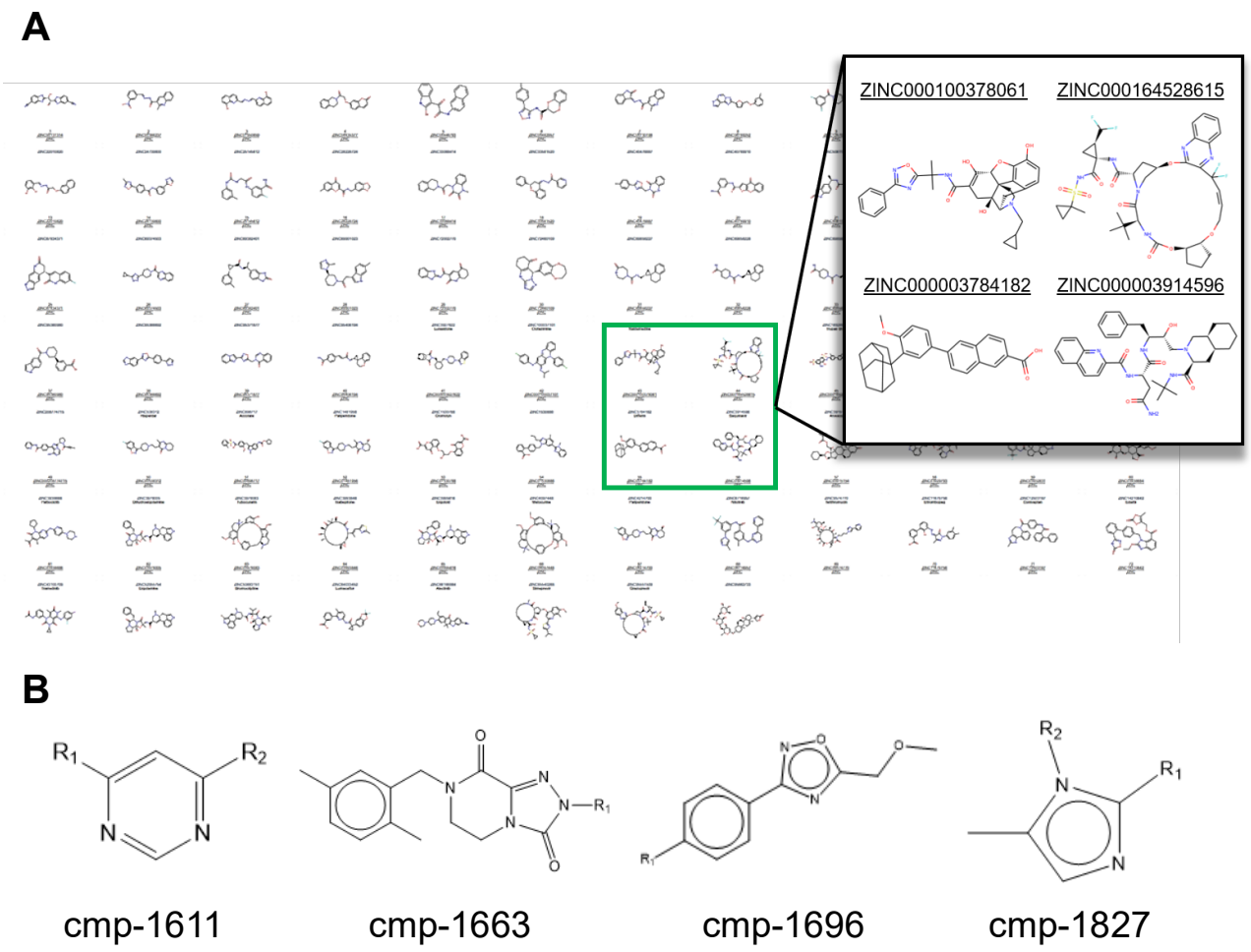


Figure 15. Structurally similar compounds derived from *in silico* docking experiment. An alternate screening method generated a list of 756 additional structurally diverse compounds predicted to bind C1r-CCP2-SP. **(A)** Of those compounds, 80 compounds exhibited favorable physicochemical properties for C1r-binding and inhibition. The magnified box displays four randomly selected compounds of the 80 to illustrate structural distinction between the screened compounds. In total, 21 of those were identified to have at least one of the same core fragments of four lead compounds. **(B)** Structural similarity determinations were based on these fragments unique to the four compounds and their respective analogues.

We selected the top seven compounds that shared structural similarities to our lead compounds from the *in silico* docking experiment based on favorable xLogP values (compounds with greater affinity for aqueous phase). Of those seven, one shared core fragments with cmp-1611, two with cmp-1663, one with cmp-1696, two with cmp-1827, and one shared core fragments with cmp-1611, cmp-1663, and cmp-1696. Additionally, three compounds are already available on the market as drugs, and are nilotinib, lumacaftor, and edarbi. The seven compounds were ordered for testing and arbitrarily labelled docking hit 001 through docking hit 007 (DH001-007).

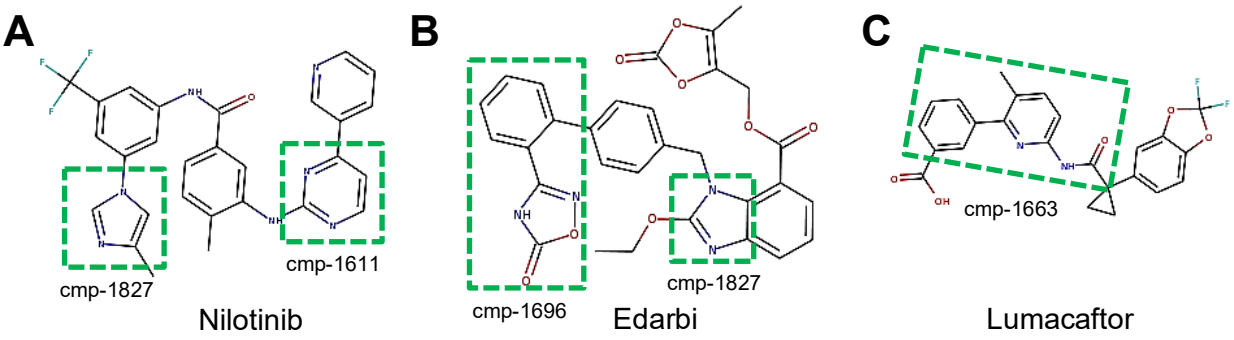


Figure 16. Structural similarity of core scaffolds within available drugs from *in silico* docking experiment. (A-C) From the seven selected compounds of the docking experiment that shared structural similarities with our four lead compounds, three are available on the market currently as drugs. Nilotinib shares core scaffolds with cmp-1611 and cmp-1827. Edarbi shares structural similarities with cmp-1696 and cmp-1827. Lumacaftor has a similar core fragment to cmp-1663.

These seven docking hits were resuspended to concentrations of either 10 mM, 50 mM, or 100 mM based on solubility in DMSO. They then underwent testing for classical complement inhibition in both enzymatic cleavage assays and ELISA-based *in vitro* assays (Figure 17). None of the compounds generated reproducible inhibition results under either experimental method, especially when compared to known nonselective serine-protease inhibitor, futhan, which functioned as a basis for comparison of effective dose-dependent inhibition.

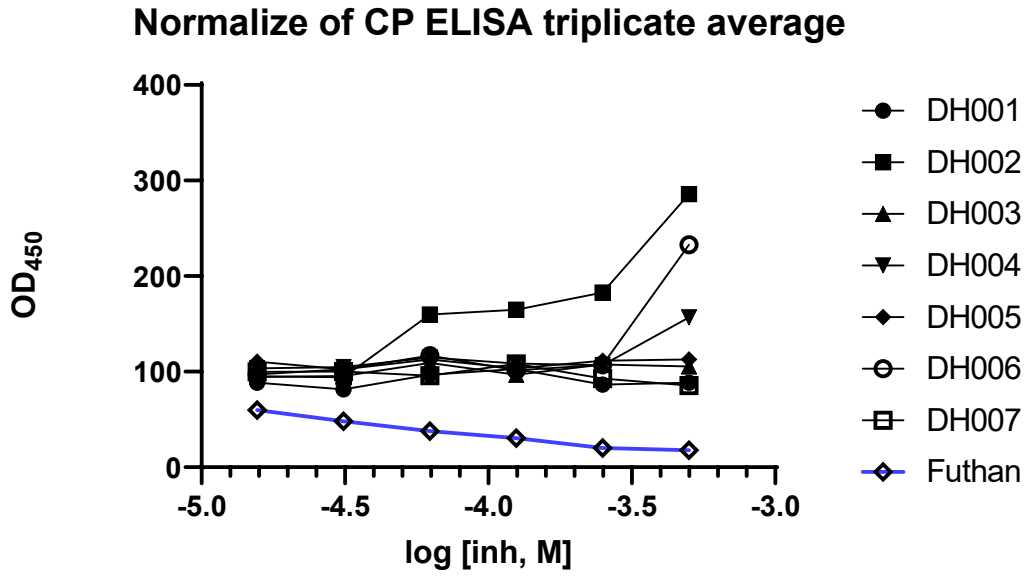


Figure 17. Activity of common core fragment compounds in ELISA. Docking hits were tested using 500 μM final concentration against normal human serum (NHS) for inhibition of C4b deposition. Despite *in silico* docking predictions near or within the C1r-CCP2-SP domain, none of the seven selected docking hits exhibited reproducible classical complement inhibition.

CHAPTER 4: CHEMINFORMATICS-BASED APPROACH

4.1 Molecular docking analysis of C1s-binding compounds

Intervention at the C1 level can be achieved outside of C1r by targeting secondary protease, C1s, whose activation is also critical for initiation of the classical complement cascade, as its only two native substrates are C2 and C4. In a parallel approach, we used a cheminformatics-based approach to identify small-molecule compounds with druglike characteristics that are predicted to bind C1s. Using the published structure of C1s, four critical sites responsible for enzymatic function and activation were established. These docking sites were determined based on the intermolecular contacts that C1s makes within the C1 complex. These structure-guided determinants were identified using published C1s crystal structures, cryo-EM, and small angle x-ray scattering studies (SAXS). The sites, i) C1q/C1s interface, ii) C1r-C1s interface, iii) C1s dimer interface, and iv) C1s scissile loop, were defined in Chemalytic's implementation of AutoDock Vina, a program designed to computationally index druglike compounds and their properties from *in silico* libraries. Utilizing the public-access ZINC database, 4.27 million compounds were theoretically docked within one of the four sites (Figure 18a).

Due to its activity and function as a serine protease, we targeted the active site within the serine protease domain of C1s for therapeutic intervention by virtual screening of small-molecule compounds from the ZINC database. Using AutoDock Vina, compounds were screened for C1s docking potential (Figure 18B). In total, approximately 150,000 drug-like compounds and FDA-approved small-molecule drugs were screened against C1s.

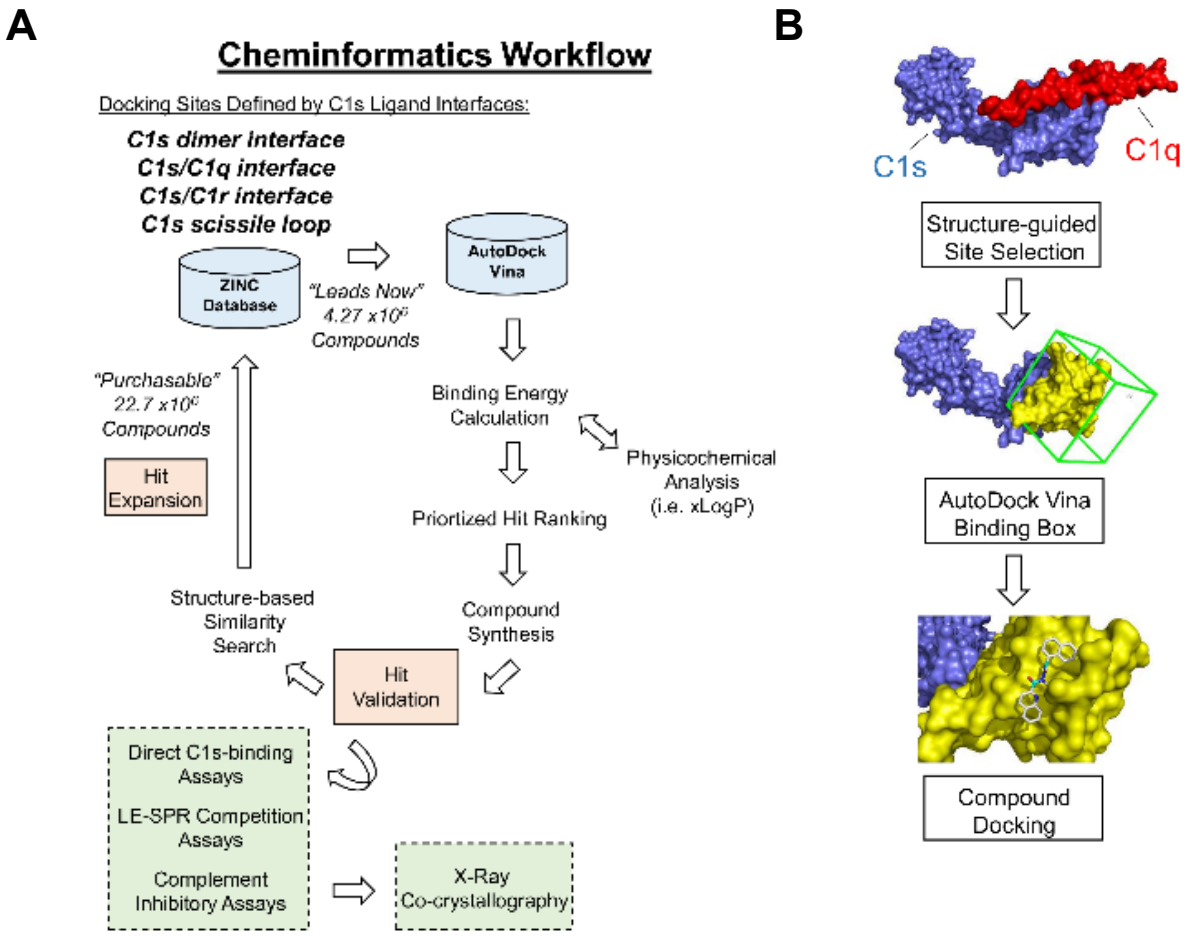


Figure 18. C1s Cheminformatics workflow and *in silico* docking schematic using AutoDock Vina binding box. (A) Critical docking sites were established using C1s structure-based design. Virtual screening of the ZINC compound database yielded compound leads that were prioritized and ranked based on binding energy and physicochemical properties to generate "hits" for *in vitro* validation assays. Compound hits will be expanded through structure-based similarity searches for additional promising compound leads. **(B)** Using published C1s crystal structure, molecular mechanisms of function, activation, and ligand interaction can be better understood for determination of critical target sites on the serine protease domain. Those sites can be defined using AutoDock Vina to virtually screen for potential compound binders.

The hits generated from *in silico* docking experiments were then prioritized based on the averaged calculated binding energies of the six most favorable compound conformations. Those with a score of 4.5 standard deviations above the average docking score were analyzed for their xLogP aqueous solubility values (Figure 19). The solubility of a compound in aqueous solution is of concern in drug development because hydrophobicity can drive toxicity as it is generally retained longer in the body. Additionally, insoluble compounds can damage SPR instrument components and give false reads [56]. Using ChemMine's hierarchical clustering software, compounds with high theoretical C1s-binding affinity, favorable physicochemical properties, and diverse chemical structures were selected for further testing.

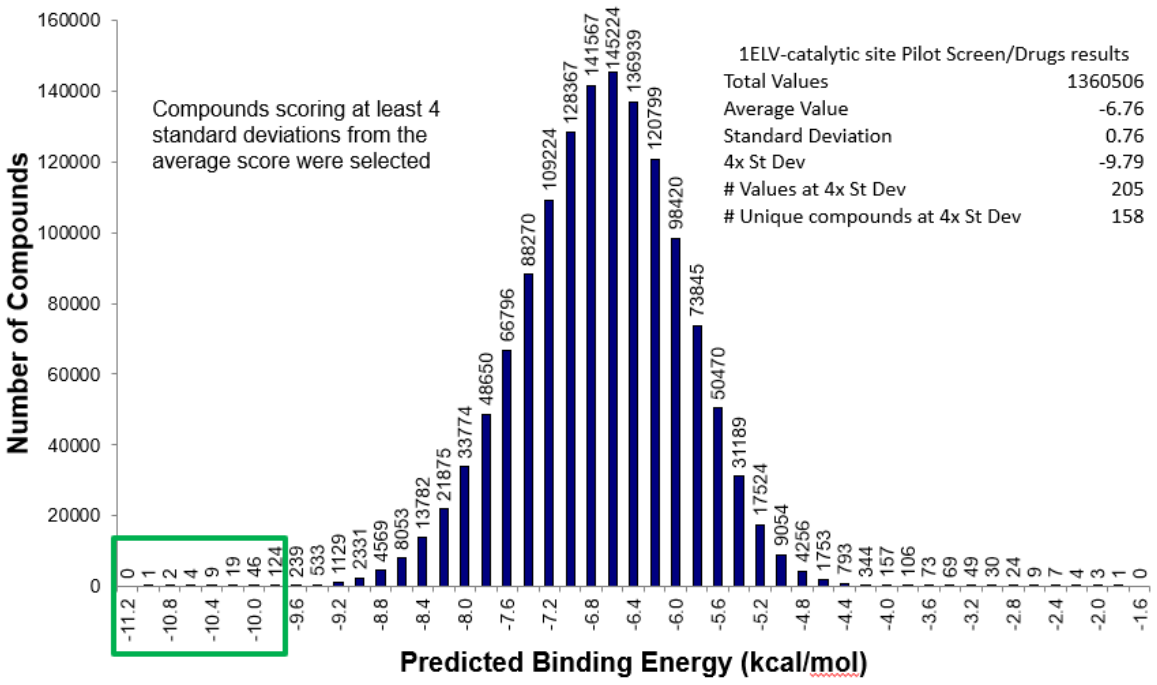


Figure 19. C1s binding box screened against ~150,000 drug-like compounds and FDA-approved small-molecule drugs from the ZINC database. Computationally-derived binding energies were averaged and prioritized. Compounds scoring 4.5 standard deviations above the average docking score were selected for further screening (green box). xLogP values were assessed and compounds with poor solubility in aqueous solution were eliminated.

4.2 C1s-binding and inhibition properties of triaged compounds

Fifty-two compounds were selected for *in vitro* validation using the SPR, as previously described for C1r. Selected compounds were commercially synthesized (ChemDiv) and obtained in larger quantities. Compounds were resuspended in DMSO to a stock concentration of 100 mM for use in SPR, they were further diluted to a concentration of 500 μM and injected over full-

length C1s immobilized on an SPR sensorchip to assess binding. Compounds that exhibited C1s binding characteristics were injected over recombinantly expressed C1s-CCP2-SP domain truncation over varying concentrations ranging from 7.8 to 500 μM to measure the dose response. In total, 28 compounds have been assessed and ranked based on binding affinity. From the prioritized compounds, compound B-5 emerged as having the greatest, reproducible binding affinity to C1s with a K_D of 5 μM to that of 38 μM in control compound, futhan. B-5 was tested for dose-dependent inhibition of the classical pathway in two-fold multivariate concentrations by measurement of C4b deposition in ELISA-based classical pathway inhibition assays and showed no reproducible inhibition (Figure 20C).

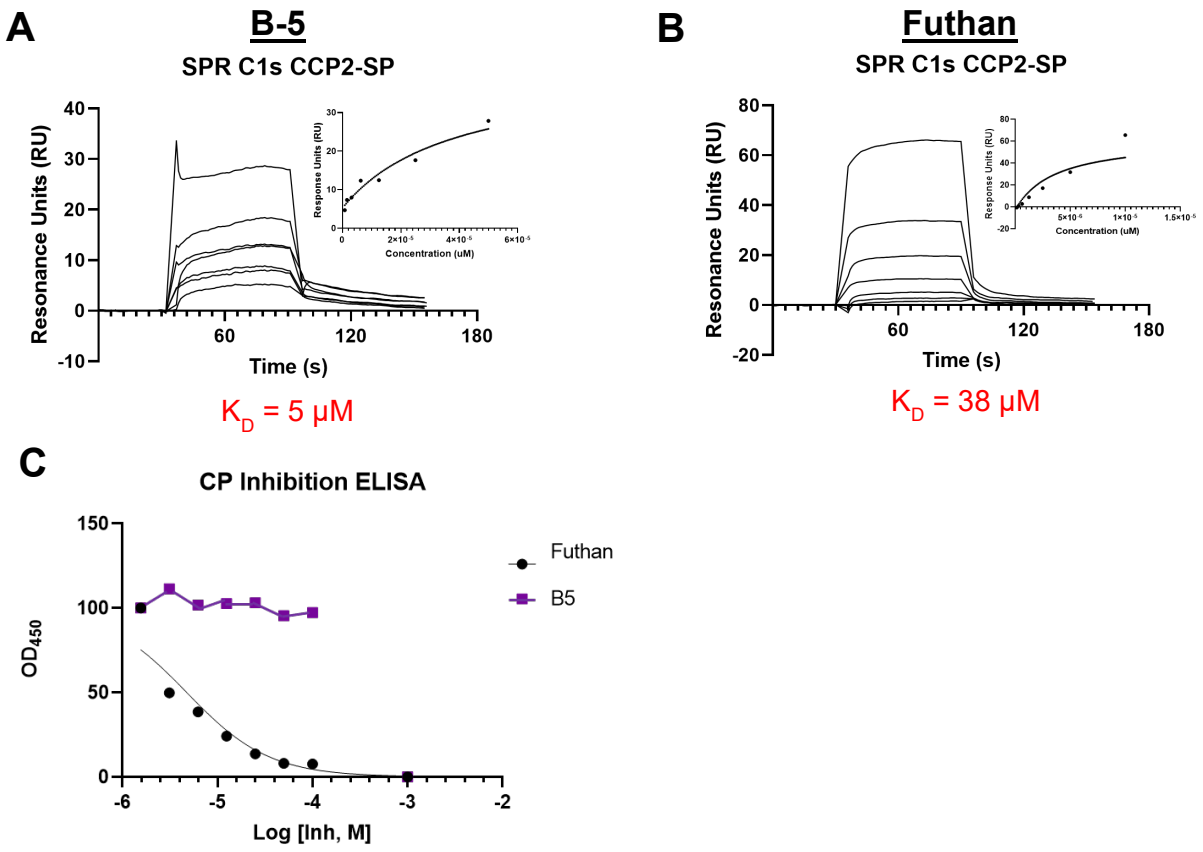


Figure 20. Dose-dependent binding of C1s-CCP2-SP by compound B-5 versus promiscuous serine protease inhibitor, futhan, does not yield inhibition of the classical pathway. (A-B) Compound B-5 and futhan were injected over immobilized C1s-CCP2-SP domain truncation and assessed for dose-dependent binding. B-5 exhibited similar binding affinity as the control, futhan, K_D = 5 µM and K_D = 38 µM, respectively. **(C)** Despite its binding capabilities, B-5 did not inhibit classical pathway activation in ELISA-based inhibition assays. All experiments were performed in triplicate.

As of yet, no small-molecule compounds from our study have been identified as potent C1s inhibitors. Despite this, C1s remains a promising therapeutic target for ameliorating classical

pathway-mediated diseases. Increased screening and evaluation of potential hit compounds is needed for future studies. We plan on expanding our compound list and continuing to assess promising leads generated by *in silico* computational docking via *in vitro* biochemical and functional validation assays.

CHAPTER 5: CRYSTALLOGRAPHY

5.1 C1r-2SP WT

To obtain empirical binding data for our small-molecule compounds in complex with C1r, we strove to produce high-resolution x-ray diffraction data from co-crystallization. In order to attain a suitable protein concentration for obtaining crystal structures, we produced 3 mg/mL of C1r-CCP2-SP domain truncation (after correcting for the extinction coefficient of 1.6). We achieved recombinant expression of C1r-CCP2-SP using BL21 *E. coli* and purified the protein using gel filtration on FPLC. After having produced sufficient concentration of protein, we concentrated our samples down to the target concentration of 3 mg/mL using 3K MWCO microcentrifuge spin filters and buffer exchanged the protein into 20 mM HBS solution. Using the wide range of crystal conditions provided by Hampton Crystal and Index Screens, we transferred both C1r-truncation and 5 mM concentration of target compound into sitting drop trays using the crystal griffon instrument for co-crystallization through compound soak.

In doing so, two conditions provided working crystals in less than 48 hours after transfer to the sitting drop trays. Select crystals were cryoprotected and sent to Argonne National Labs for x-ray diffraction data collection using beamline 22-ID or 22-BM of the Advanced Photon Source. X-ray diffraction data were collected by shooting the crystal with an x-ray laser over 180° of the crystal surfaces. Once x-ray diffraction data were collected, we utilized the published 2.8Å resolution crystal structures of C1r-CCP2-SP (1GPZ) to solve our structure of 1.8Å resolution by accounting for rotamer analyses, bond constraints, and electron densities, which indicated the presence or absence of structural components, among others using WinCoot for crystallographic structural imaging and manipulation in conjunction with Phenix crystallographic

molecular structure determination software. After completion of the solved crystal structure using Phenix, an unexplained electron density remained within the substrate specificity S1 site of the serine protease domain. Though we had anticipated that our compounds would dock within this site, further analysis revealed that rather than binding our compound in complex with C1r, the flexible rotamer of arginine-463 of the serine protease domain blocked availability of the active site (Figure 21A). When in tight crystal complex with other C1r-CCP2-SP proteins within the asymmetric unit, the arginine-463 rotamer invariably folded within the S1 site, blocking any compound docking. Trypsin-like serine proteases, such as C1r and C1s, are often characterized by a negatively charged aspartate located at the base of the S1 pocket [9]. This negative charge drives specificity for cleavage of positively charged amino acids. In tight complex with other C1r-CCP2-SP proteins, it is likely that positively charged arginine-463 is driven to the S1 site by the negatively charged aspartate. To overcome that unforeseen obstacle, we decided to create a C1r-CCP2-SP domain truncation mutant in which the arginine is replaced with a glutamine to eliminate the active site blockage (Figure 21B).

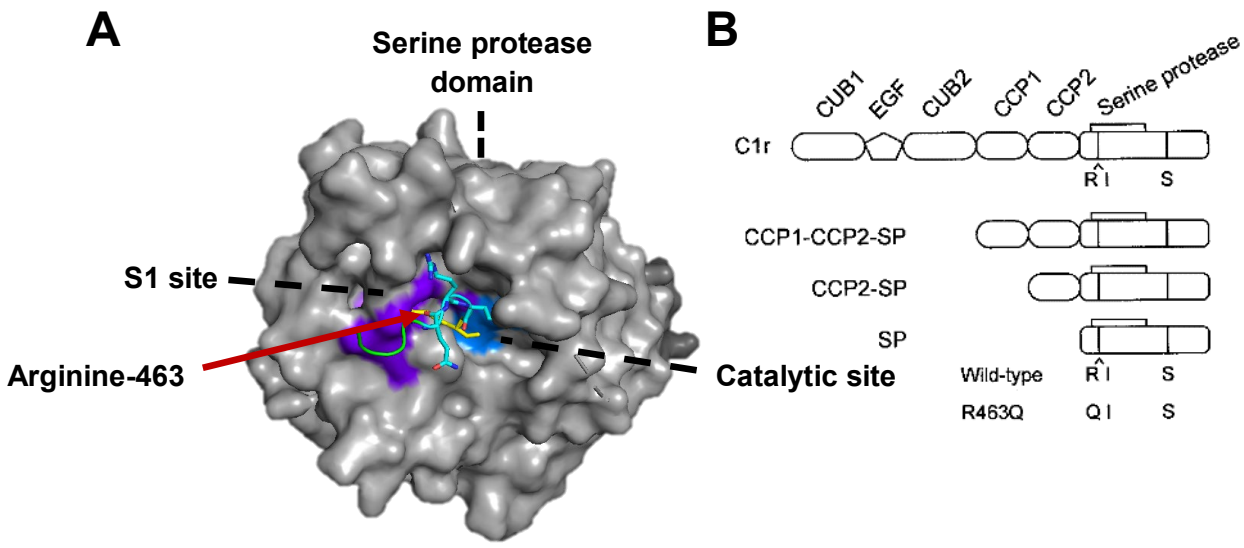


Figure 21. Mutating arginine-463 for glutamate allows for accessibility of the S1 site in C1r-CCP2-SP domain truncation crystal. (A) The arginine side chain (yellow residue) of the serine protease domain of neighboring C1r folds into the S1 site (purple) of WT C1r-CCP2-SP (gray). The figure depicts a portion of neighboring C1r-2SP protein (green, cyan, and yellow residues) interacting with C1r-CCP2-SP. When in tight crystal formation, this action obstructs accessibility to the S1 or catalytic site (marine blue) by small-molecule compounds. **(B)** To address active site availability, arginine-463 is mutated to glutamine (Q) to allow for small-molecule binding.

5.2 C1r-2SP R463Q

To address the obstacle of the arginine-463 side chain of neighboring C1r-CCP2-SP from blocking the S1 site of C1r, we created a C1r-CCP2-SP domain truncation mutant in which the arginine was substituted with a glutamine. This protein structure was recombinantly expressed in an identical manner to C1r-CCP2-SP WT by BL21 *E. coli* and concentrated to 3 mg/mL. The

protein was transferred into 20 mM HBS and soaked with our four lead compounds, yielding two successful crystal conditions containing three crystal structures that provided 1.8 Å high-resolution x-ray diffraction data. Using published C1r-CCP2-SP R463Q mutant PDB model (1GPZ), with Wincoot and Phenix software, all three structures were solved, though none contained our compounds.

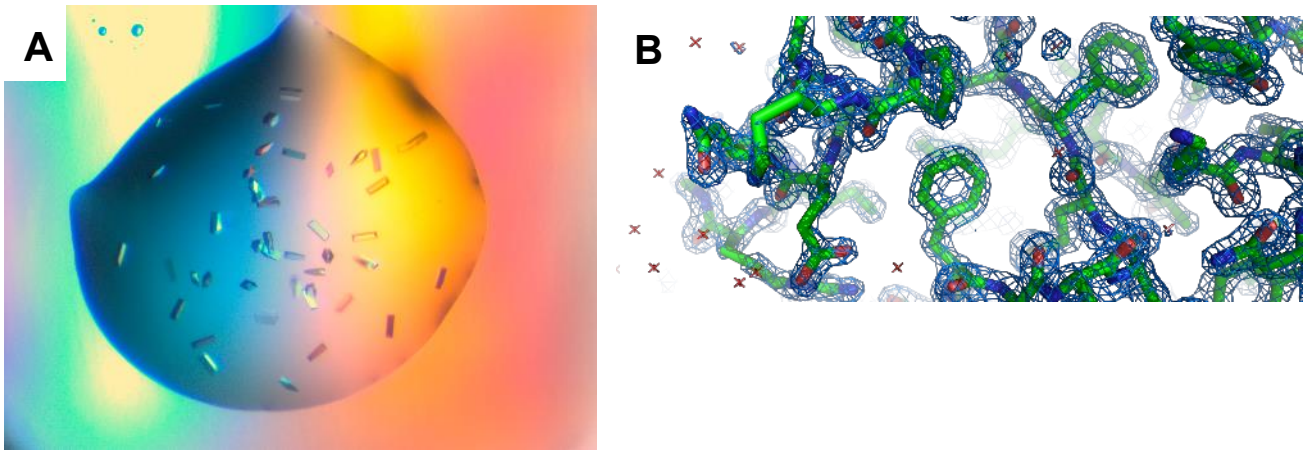


Figure 22. C1r-CCP2-SP domain truncation mutant suspended in crystal screen reagent drop and its solved molecular crystal structure. (A) C1r-CCP2-SP R463Q mutants were crystallized in 0.2 M sodium acetate trihydrate, 0.1 M Tris hydrochloride pH 8.5, and 30% w/v polyethylene glycol 4,000. These crystals produce a birefringent effect upon interaction with polarized light due to two distinct indices of refraction. **(B)** Electron density maps generated by C1r PDB model 1GPZ in Phenix and Wincoot allowed for molecular structural determination of our crystals. Upon completion of the solved structure, the electron density maps did not reveal the presence of our lead compounds in complex with C1r-CCP2-SP R463Q mutants.

Data collection and refinement	C1r-CCP2-SP$_{(307-705)}$
Data collection	
Space group	P 2 2_1 2_1
Cell dimensions	
a, b, c, Å	37.87, 91.41, 91.95
α, β, γ, °	90.00, 90.00, 90.00
Resolution, Å	41.1 – 1.70 (1.76 – 1.70)
R_{pim} [a]	0.078 (0.383)
$I/\sigma I$ [b]	18.14 (2.14)
Completeness, %	99.0 (99.6)
Redundancy	5.7 (5.3)
Refinement	
Resolution, Å	41.0 – 1.70
No. reflections	35,680
R_{work}/R_{free} [c]	19.58 / 21.99
No. non-hydrogen atoms	2,732
Protein	2,476
Water	256
B-factors	
Protein	23.77
Water	31.90
RMSD [d]	
Bond lengths, Å	0.010
Bond angles, °	1.123

Table 1. C1r-CCP2-SP R463Q data collection and refinement statistics (molecular replacement). Values in parentheses are represent the highest-resolution shell. The crystal structure of C1r-CCP2-SP R463Q was solved using Phaser-MR using the previously published structure of C1r-CCP2-SP (1GPZ). The final model was obtained using WinCoot-CCP4 and refinement in Phenix 1.16-3549. Models were validated using Molprobity and the structural representations were prepared using PyMOL.

[a] R_{pim} is the measure of precision of the averaged intensities and amplitudes.

[b] $I/\sigma I$ is the precision indicator for the individual observations, where σ is error propagation.

[c] $R = \sum_h | F_o(h) - F_c(h) | / \sum_h | F_o |$. R_{work} and R_{free} were calculated from the working and test reflection sets, respectively. The test set constituted 10% of the total reflections not used in refinement.

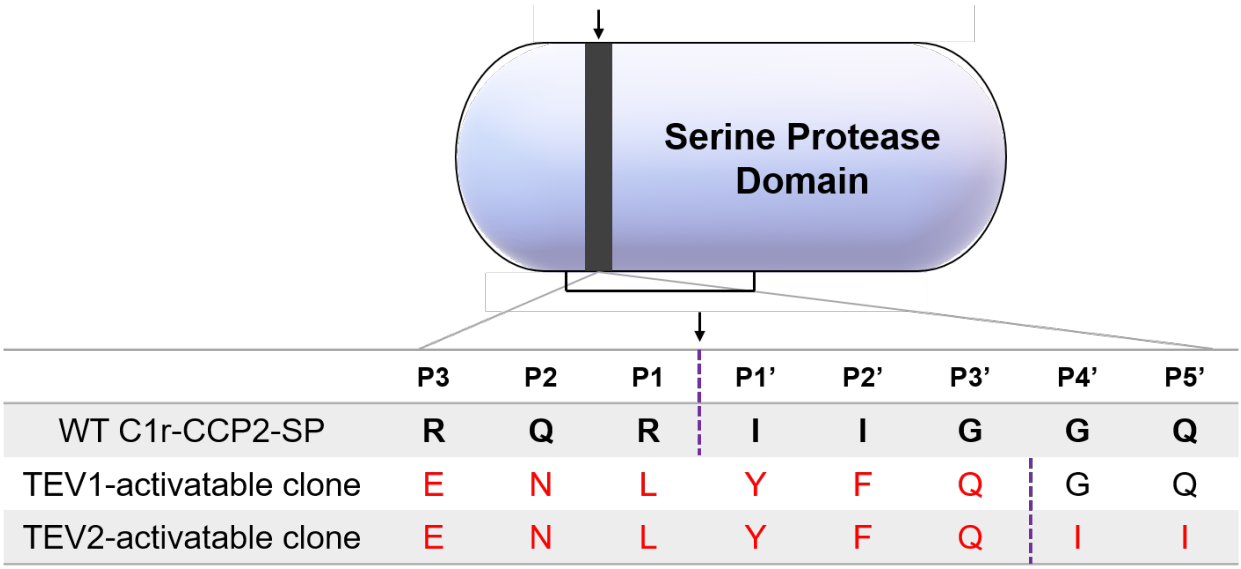
[d] RMSD is the root-mean-square deviation of atomic particles that accounts for the measure of average distance between particles.

5.3 C1r-2SP-S654A

Another issue arising from the crystallization of C1r-CCP2-SP was that C1r was autoactivating, even in the presence of inhibitor. In order to disrupt autoactivation of C1r, we mutated serine-654 of the catalytic triad to an alanine to stabilize the zymogen form of C1r-CCP2-SP. C1r-CCP2-SP S654A mutant was recombinantly expressed by BL21 *E. coli* and crystallized in solution with inhibitor. X-ray diffraction data obtained from the S654A mutant crystals showed that despite mutation of a key residue within the active site, C1r had undergone activation. We speculate that bacterial proteases from *E. coli* activate C1r during synthesis or purification of C1r-2SP.

5.4 C1r-2SP TEV-activatable clone

To control activation of C1r, we designed clones that can only be activated by tobacco etch virus (TEV) directed to a TEV-specific DNA sequence insert in transformed pT7HMT plasmid into DH5α *E. coli.* To achieve this end, we mutated the conserved RQRIIG residues of wild-type human C1r-CCP2-SP to the TEV-activatable sequence, ENLYFQ (Figure 23). We hypothesize that this transformation will inhibit C1r activation and allow for binding by an inhibitor to the active site of C1r. C1r-TEV-activatable mutants were recombinantly expressed by BL21 *E. coli,* though no studies have yet taken place with these mutants.



	P3	P2	P1	P1'	P2'	P3'	P4'	P5'
WT C1r-CCP2-SP	**R**	**Q**	**R**	**I**	**I**	**G**	**G**	**Q**
TEV1-activatable clone	E	N	L	Y	F	Q	G	Q
TEV2-activatable clone	E	N	L	Y	F	Q	I	I

Figure 23. Site-directed mutagenesis to control C1r activation. The S3-S5' sites denote specific amino acid residues that comprise the molecular determinants for C1r-substrate specificity. The black arrows indicate the disulfide bridge between arginine and isoleucine that must be cleaved in order to achieve C1r activation. The amino acids in bold represent those present in WT C1r-CCP2-SP, whereas those in red have been exchanged for *E. coli*-optimized tobacco etch virus (TEV) sequences. By withholding the TEV protease, C1r is expected to remain in its zymogen form.

CHAPTER 6: DISCUSSION

Advancements in scientific research and medical technologies that are rapidly expanding our understanding of molecular disease mechanisms and triggers comes at no better time than now as our US population is aging. Within the next 40 years, the number of Americans aged 65 and older will more than double, placing increased weight on developing therapeutics directed towards age-related degenerative disorders. One such example is Alzheimer's disease, which is significantly burdening our aging population, as well as government resources and healthcare system. It is currently the sixth leading cause of death in the United States, and its impact is only expected to escalate [71]. Neurodegenerative inflammatory diseases, such as Alzheimer's disease, have complex pathogenic mechanisms, and complement's involvement in exacerbating and/or driving pathogenesis is becoming increasingly evident. Given the complexities of each disease, there will obviously be no "cure-all" drug and the need for specific disease-tailored therapeutics, especially in the anticomplement drug field, serves as the catalyst for many studies such as this.

The successes of complement-directed drugs, eculizumab and C1-INH, on the treatment of a growing list of diseases has put the spotlight on anticomplement therapeutics. Currently, there are over 40 different anti-complement drugs in various stages of clinical development [49]. The majority of these drugs focus on targeting the complement pathway convertases. However, complement therapeutics directed at the C3 and C5 level overlook pathway-specific disease-initiation mechanisms in a wide spectrum of complement-mediated diseases, such as Alzheimer's disease. Moreover, targeting a specific activation pathway presents the capacity for continued immune surveillance by the remaining intact pathways.

Despite increased interest in the treatment of complement-mediated disorders, there remains a void in drugs targeting the initiating proteases of the classical and lectin pathways. The

classical complement pathway has been implicated in a growing disease inventory due to its diverse assortment of pathway activators. Historically, complement's oldest known role was as a support function in antibody-mediated bacterial clearance. Indeed, the classical pathway is frequently activated by immune complex recognition and binding by C1q to the surface membrane of foreign invaders, cellular debris, and/or host cells, such is the case with heparin-induced thrombocytopenia, neuromyelitis optica, bullous pemphigoid, and ischemia-reperfusion organ injuries [20,72–74]. However, classical pathway activation can occur by C1q attachment to over 100 diverse ligand activators, such as amyloid-β plaques and neurofibrillary tangles, as implicated in Alzheimer's disease [75].

At present, there is a paucity of classical pathway-specific drugs, with only three currently available on the market. Recombinant and native preparations of C1-INH (Cinryze/Berinert/Ruconest) have been approved by the FDA for treatment of hereditary angioedema, and they are also now being evaluated more broadly for treatment classical and lectin pathway-mediated pathologies, such as acute antibody-mediated rejection following allogeneic organ transplantation and septicemia/SIRS [28]. Anti-C1q antibody ANX005 (Annexon Biosciences) and anti-C1s antibody TNT009/BIV009 (True North Therapeutics) are presently in phase I and phase III clinical trials for treatment of Guillain Barré syndrome and cold agglutinin disease, respectively [49]. Inhibition of the classical pathway cascade through C1-directed drugs is showing promise for the treatment of a broad range of classical pathway-mediated diseases in the current clinical landscape, making the pathway initiating serine proteases, C1r and C1s, favorable therapeutic targets.

Still, there are many challenges facing the development of anticomplement drugs, such as dosage, tissue penetrance, and specificity. Antibody development against complement

components is the most prevalent intervention strategy, but its use has highlighted obstacles in complement therapeutics. Due to the abundance of complement proteins in plasma and the high rate of turnover, delivering sufficient concentrations of such a large inhibitor as an antibody is problematic, with breakthrough thrombosis a possible consequence, as reported with eculizumab treatment [76]. Small-molecule inhibitors can overcome this pitfall through oral bioavailability, which is not possible with anti-complement antibodies. Repeat dosing by orally bioavailable small molecules allows for treatment of chronic conditions [50]. Small-molecule inhibitors can also be rapidly cleared from the body once dosing ceases for full restoration of complement activity for immune function. Small-molecule drugs also have an advantage over biologics by being able to diffuse to sites of tissue damage and inflammation. This limitation of biologics is exacerbated in organs protected by tight membrane barriers, such as the blood-brain-barrier [77]. In immune privileged disease sites, such as the brain or eye in neurodegenerative and ocular diseases, barrier dysfunction is often observed resulting in leaky cell junctions, easing tissue penetrance by small-molecule inhibitors [78]. Lack of specificity with small-molecules can be overcome with structure-based drug design where a full appreciation of the binding modes, protein interactions, and specific surface features of the target complement protein can allow for optimization of selectivity in small-molecules [79]. Major breakthroughs in small-molecule drug development have recently emerged with the development of small-molecule antagonists against factor B and factor D of the alternative pathway [55,66]. The development of orally bioavailable small-molecule inhibitors has the added benefits of increasing patient compliance and lowering costs associated with these therapies.

In our study, we aimed to target the initiating proteases, C1r and C1s, of the classical pathway of complement to halt pathway activation at the level of C1. Classical pathway

initiation occurs upon C1q attachment to a target site, which sets off the autoactivation of the chymotrypsin-like serine protease, C1r. C1r then functions to proteolytically cleave the subsequent C1 protease, C1s. The activation of these two proteases kicks off the classical pathway cascade, rendering C1r and C1s the most upstream enzymatic targets of the pathway. Through the identification and development of small-molecule inhibitors of these proteases, we hope to specifically inhibit activation of the classical pathway and expand the anticomplement-therapeutic toolkit.

To develop small-molecule inhibitors, we employed two structure-based, orthogonal approaches to compound screening. Using fragment-based drug discovery, we screened compound libraries for fragments that can be optimized for highly competitive binding interactions with C1r and high levels of specificity. Our fragment search and screening produced four structurally distinct small-molecule inhibitors of C1r: cmp-1611, cmp-1663, cmp-1696, and cmp-1827 that all bind near or within the catalytic cleft of the serine protease domain. Optimization strategies to augment identified fragments into compound leads has seen great success in the development of potent, drug-like small-molecule inhibitors through the process of fragment linking, growing, and/or merging, all of which rely on careful structure-based design [80]. The success of factor D small-molecule inhibitor development is a prime example of the potential of small-molecules in enzymatic complement inhibition [55]. In a parallel approach, we utilized cheminformatics to virtually screen compound libraries for drug-like compounds predicted to bind within specific target sites on the C1s serine protease. Compound B-5, derived from screening drug-like compounds of the “Leads Now” subset of the ZINC database, showed promise for its ability to dose-dependently bind C1s. Yet, its binding affinity was not representative of its ability to block C1s activity. As of yet, no inhibitory compounds have been

identified for C1s, necessitating continued efforts in the evaluation of compound leads. High-resolution crystal structures of our recombinantly expressed and renatured C1r/C1s domain truncations and domain mutants will significantly advance and promote the success of this study. The molecular nuances of protein-protein interactions and critical surface structures from our crystallographic efforts were extremely helpful for guiding the development of our small-molecule antagonists.

Our work has produced four structurally distinct, small-molecule compounds that bind the active site of the initiator protease of the classical pathway of complement, C1r. Each compound will, in a dose-dependent manner, block C1r activation, thereby preventing classical pathway activation. In addition, each compound has favorable physicochemical properties for independent or combinatorial optimization strategies. It is our hope that continued efforts in co-crystallization and optimization of these compounds will allow for the advancement of highly specific, high-affinity small-molecule inhibitors of C1r and C1s. The development of this type of inhibitor targeting the activation of C1 could provide valuable and novel insights for the mechanisms of C1 inhibition, as well as provide potential treatment options for diseases mediated by the aberrant activation of the classical pathway of the complement system.

REFERENCES

1. Nesargikar, P. N., Spiller, B. & Chavez, R. The Complement System: History, Pathways, Cascade and Inhibitors. *Eur. J. Microbiol. Immunol.* **2**, 103–111 (2012).

2. Kaufmann, S. H. E. Immunology's foundation: The 100-year anniversary of the Nobel Prize to Paul Ehrlich and Elie Metchnikoff. *Nat. Immunol.* **9**, 705–712 (2008).

3. Ricklin, D., Reis, E. S. & Lambris, J. D. Complement in disease: a defence system turning offensive. *Nat. Rev. Nephrol.* **12**, 383–401 (2016).

4. Merle, N. S., Church, S. E., Fremeaux-Bacchi, V. & Roumenina, L. T. Complement system part I - molecular mechanisms of activation and regulation. *Frontiers in Immunology* **6**, (2015).

5. Forneris, F. *et al.* Structures of C3b in Complex with Factors B and D Give Insight into Complement Convertase Formation. *Science (80-.).* **330**, 1816–1820 (2010).

6. Klop, B. *et al.* Differential complement activation pathways promote C3b deposition on native and acetylated LDL thereby inducing lipoprotein binding to the complement receptor 1. *J. Biol. Chem.* **289**, 35421–35430 (2014).

7. Gál, P., Barna, L., Kocsis, A. & Závodszky, P. Serine proteases of the classical and lectin pathways: Similarities and differences. *Immunobiology* **212**, 267–277 (2007).

8. Gaboriaud, C., Frachet, P., Thielens, N. M., Arlaud, G. J. & Quax, P. The human C1q globular domain: structure and recognition of non-immune self ligands. (2012). doi:10.3389/fimmu.2011.00092

9. Kardos, J. *et al.* The Role of the Individual Domains in the Structure and Function of the Catalytic Region of a Modular Serine Protease, C1r. *J. Immunol.* **167**, 5202–5208 (2001).

10. Kardos, J. *et al.* Revisiting the mechanism of the autoactivation of the complement protease C1r in the C1 complex: Structure of the active catalytic region of C1r. *Mol. Immunol.* (2008). doi:10.1016/j.molimm.2007.09.031

11. Almitairi, J. O. M. *et al.* Structure of the C1r-C1s interaction of the C1 complex of complement activation. *PNAS* **115**, 768–773 (2018).

12. Mortensen, S. A. *et al.* Structure and activation of C1, the complex initiating the classical pathway of the complement cascade. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 986–991 (2017).

13. Degn, S. E. *et al.* Complement activation by ligand-driven juxtaposition of discrete pattern recognition complexes. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 13445–13450 (2014).

14. Bohlson, S. S., Garred, P., Kemper, C. & Tenner, A. J. Complement Nomenclature-Deconvoluted. *Front. Immunol.* **10**, 1308 (2019).

15. Holers, V. M. Complement and Its Receptors: New Insights into Human Disease. *Annu. Rev. Immunol.* **32**, 433–459 (2014).

16. Morgan, B. P. Complement in the pathogenesis of Alzheimer's disease. *Semin.*

Immunopathol. **40**, 113–124 (2018).

17. Strang, C. J. *et al.* Angioedema Induced by a Peptide Derived from Complement Component C2. *J Exp Med* **168**, 1685–1698 (1988).

18. Ricklin, D. & Lambris, J. D. Complement in Immune and Inflammatory Disorders: Pathophysiological Mechanisms. *J Immunol Ref.* **190**, 3831–3838 (2019).

19. Saadoun, S. & Papadopoulos, M. C. Role of membrane complement regulators in neuromyelitis optica. *Mult. Scler. J.* **21**, 1644–1654 (2015).

20. Freire, P. C. *et al. Specific Inhibition of the Classical Complement Pathway Prevents C3 Deposition along the Dermal-Epidermal Junction in Bullous Pemphigoid*. *Journal of Investigative Dermatology* (Society for Investigative Dermatology, 2019). doi:10.1016/j.jid.2019.04.025

21. Ricklin, D., Reis, E. S., Mastellos, D. C., Gros, P. & Lambris, J. D. Complement component C3 - The 'Swiss Army Knife' of innate immunity and host defense. *Immunol. Rev.* **274**, 33–58 (2016).

22. Carroll, M. C. & Isenman, D. E. Regulation of Humoral Immunity by Complement. *Immunity* **37**, 199–207 (2012).

23. Khandelwal, S. *et al.* Polyreactive IgM initiates complement activation by PF4/heparin complexes through the classical pathway. *Blood* **132**, 2431–2440 (2018).

24. Kalogeris, T., Baines, C. P., Krenz, M. & Korthuis, R. J. Cell Biology of Ischemia/Reperfusion Injury. *Int Rev Cell Mol Biol* **298**, 229–317 (2012).

25. Castellano, G. *et al.* Therapeutic targeting of classical and lectin pathways of Complement protects from ischemia-reperfusion induced renal damage inhibiting the activation of dendritic cells. *Abstr. / Mol. Immunol.* **46**, 2818–2871 (2009).

26. Kwan, W. hong, van der Touw, W., Paz-Artal, E., Li, M. O. & Heeger, P. S. Signaling through C5a receptor and C3a receptor diminishes function of murine natural regulatory T cells. *J. Exp. Med.* **210**, 257–268 (2013).

27. Distelmaier, K. *et al.* Local complement activation triggers neutrophil recruitment to the site of thrombus formation in acute myocardial infarction. *Thromb. Haemost.* **102**, 564–572 (2009).

28. Sungurtekin, H., Balcý, C. & Sungurtekin, U. Circulating complement proteins for differentiation of SIRS and sepsis. *Crit. Care* **7**, P031 (2003).

29. Merle, N. S., Noe, R., Halbwachs-Mecarelli, L., Fremeaux-Bacchi, V. & Roumenina, L. T. Complement system part II: Role in immunity. *Frontiers in Immunology* **6**, (2015).

30. Al-Ani, F., Chin-Yee, I. & Lazo-Langner, A. Eculizumab in the management of paroxysmal nocturnal hemoglobinuria: Patient selection and special considerations. *Therapeutics and Clinical Risk Management* **12**, 1161–1170 (2016).

31. Waters, A. M. & Licht, C. AHUS caused by complement dysregulation: New therapies on the horizon. *Pediatric Nephrology* **26**, 41–57 (2011).

32. Loveless, S. *et al.* Tissue microarray methodology identifies complement pathway activation and dysregulation in progressive multiple sclerosis. *Brain Pathol.* **28**, 507–520 (2018).

33. Sun, M. *et al.* Schizophrenia risk from complex variation of complement component 4. *Nature* **61**, 5985–5991 (2016).

34. Depboylu, C., Schäfer, M. K.-H., Arias-Carrión, O. & Oertel, W. H. Possible Involvement of Complement Factor C1q in the Clearance of Extracellular Neuromelanin from the Substantia Nigra in Parkinson Disease. *J Neuropathol Exp Neurol* **70**, 125–132 (2011).

35. Wall, S. *et al.* Complement is Activated via the Classical Pathway (CP) in Patients with Gastrointestinal Graft Versus Host Disease (GI GVHD). *Biol. Blood Marrow Transplant.* **24**, S182–S183 (2018).

36. Wu, T. *et al.* Complement C3 Is Activated in Human AD Brain and Is Required for Neurodegeneration in Mouse Models of Amyloidosis and Tauopathy. *Cell Rep.* (2019). doi:10.1016/j.celrep.2019.07.060

37. Shen, Y. *et al.* Complement activation by neurofibrillary tangles in Alzheimer's disease. *Neurosci. Lett.* **305**, 165–168 (2001).

38. Lian, H. *et al.* Neurobiology of Disease Astrocyte-Microglia Cross Talk through Complement Activation Modulates Amyloid Pathology in Mouse Models of Alzheimer's Disease. *Neurobiol. Dis.* **36**, 577–589 (2016).

39. Litvinchuk, A. *et al.* Complement C3aR Inactivation Attenuates Tau Pathology and Reverses an Immune Network Deregulated in Tauopathy Models and Alzheimer's Disease. *Neuron* **100**, (2018).

40. Liddelow, S. A. *et al.* Neurotoxic reactive astrocytes are induced by activated microglia. *Nature* **541**, 481–487 (2017).

41. Fonseca, M. I., Zhou, J., Botto, M. & Tenner, A. J. Absence of C1q leads to less neuropathology in transgenic mouse models of Alzheimer's disease. *J. Neurosci.* **24**, 6457–6465 (2004).

42. Shi, Q. *et al.* Complement C3 deficiency protects against neurodegeneration in aged plaque-rich APP/PS1 mice. *Sci. Transl. Med.* **9**, (2017).

43. Rubin, R. D., Watson, P. D., Duff, M. C. & Cohen, N. J. The role of the hippocampus in flexible cognition and social behavior. *Front. Hum. Neurosci.* **8**, 1–15 (2014).

44. Yin, C. *et al.* ApoE attenuates unresolvable inflammation by complex formation with activated C1q. *Nat. Med.* **25**, 496–506 (2019).

45. Hong, S. *et al.* Complement and microglia mediate early synapse loss in Alzheimer mouse models. *Science (80-.).* **352**, 712–716 (2016).

46. Thurman, J. M. & Yapa, R. Complement therapeutics in autoimmune disease. *Frontiers in Immunology* **10**, (2019).

47. Harris, C. L. Expanding horizons in complement drug discovery: challenges and emerging

strategies. *Semin. Immunopathol.* **40**, 125–140 (2018).

48. Morgan, B. P. Complement in the pathogenesis of Alzheimer's disease. *Semin. Immunopathol.* **40**, 113–124 (2018).

49. Mastellos, D. C., Ricklin, D. & Lambris, J. D. Clinical promise of next-generation complement therapeutics. *Nat. Rev. Drug Discov.* **18**, 707–729 (2019).

50. Ricklin, D., Mastellos, D. C., Reis, E. S. & Lambris, J. D. The renaissance of complement therapeutics. *Nature Reviews Nephrology* **14**, 26–47 (2017).

51. Harris, C. L., Pouw, R. B., Kavanagh, D., Sun, R. & Ricklin, D. Developments in anti-complement therapy; from disease to clinical trial. *Mol. Immunol.* (2018). doi:10.1016/j.molimm.2018.06.008

52. Fraser, D. A., Laust, A. K., Nelson, E. L. & Tenner, A. J. C1q Differentially Modulates Phagocytosis and Cytokine Responses during Ingestion of Apoptotic Cells by Human Monocytes, Macrophages, and Dendritic Cells. *J. Immunol.* **183**, 6175–6185 (2009).

53. Leffler, J., Bengtsson, A. A. & Blom, A. M. The complement system in systemic lupus erythematosus: An update. *Annals of the Rheumatic Diseases* **73**, 1601–1606 (2014).

54. Drag, M. & Salvesen, G. S. Emerging principles in protease-based drug discovery. *Nature Reviews Drug Discovery* **9**, 690–701 (2010).

55. Maibaum, J. *et al.* Small-molecule factor D inhibitors targeting the alternative complement pathway. *Nat. Chem. Biol.* **12**, 1105–1110 (2016).

56. Lamoree, B. & Hubbard, R. E. Current perspectives in fragment-based lead discovery (FBLD). *Essays Biochem.* **61**, 453–464 (2017).

57. Giannetti, A. M., Koch, B. D. & Browner, M. F. Surface Plasmon Resonance Based Assay for the Detection and Characterization of Promiscuous Inhibitors. *J. Med. Chem.* **51**, 574–580 (2008).

58. Watkins, A. M. & Arora, P. S. Structure-based inhibition of protein - protein interactions. *Eur. J. Med. Chem.* **94**, 480–488 (2015).

59. Schwertz, H. *et al.* Serine protease inhibitor nafamostat given before reperfusion reduces inflammatory myocardial injury by complement and neutrophil inhibition. *J. Cardiovasc. Pharmacol.* **52**, 151–160 (2008).

60. Giannetti, A. M. From eperimental design to validated hits: A comprehensive walk-through of fragment lead identification using surface plasmon resonance. in *Methods in Enzymology* (2011). doi:10.1016/B978-0-12-381274-2.00008-X

61. Erlanson, D. A., Fesik, S. W., Hubbard, R. E., Jahnke, W. & Jhoti, H. Twenty years on: The impact of fragments on drug discovery. *Nature Reviews Drug Discovery* **15**, 605–619 (2016).

62. Ruiz-G Omez, G. *et al.* Structure-Activity Relationships for Substrate-Based Inhibitors of Human Complement Factor B. *J. Med. Chem* **52**, 6042–6052 (2009).

63. Hintze, B. J., Lewis, S. M., Richardson, J. S. & Richardson, D. C. Molprobity's ultimate rotamer-library distributions for model validation.: EBSCOhost. *Proteins* **84**, 1177–1189 (2016).

64. Keseru, G. M. *et al.* Design Principles for Fragment Libraries: Maximizing the Value of Learnings from Pharma Fragment-Based Drug Discovery (FBDD) Programs for Use in Academia. *Journal of Medicinal Chemistry* **59**, 8189–8206 (2016).

65. Baker, M. Fragment-based lead discovery grows up. *Nat. Rev. Drug Discov.* **12**, 5–7 (2013).

66. Schubart, A. *et al.* Small-molecule factor B inhibitor for the treatment of complement-mediated diseases. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 7926–7931 (2019).

67. Morris, G. M. *et al.* Software news and updates AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **30**, 2785–2791 (2009).

68. Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2126–2132 (2004).

69. Budayova-Spano, M. *et al.* The crystal structure of the zymogen catalytic domain of complement protease C1r reveals that a disruptive mechanical stress is required to trigger activation of the C1 complex. *EMBO J.* **21**, 231–239 (2002).

70. Budayova-Spano, M. *et al.* Monomeric structures of the zymogen and active catalytic domain of complement protease C1r: Further insights into the C1 activation mechanism. *Structure* (2002). doi:10.1016/S0969-2126(02)00881-X

71. Alzheimer's Facts and Figures Report | Alzheimer's Association. Available at: https://www.alz.org/alzheimers-dementia/facts-figures?utm_source=google&utm_medium=paidsearch&utm_campaign=google_grants&utm_content=alzheimers&gclid=EAIaIQobChMI1-Dd--O_5QIVrf_jBx1RmgmqEAAYASAAEgKt6fD_BwE. (Accessed: 22nd November 2019)

72. Arepally, G. M. Review Series Heparin-induced thrombocytopenia. *Blood* **129**, 2864–2872 (2017).

73. Soltys, J. *et al.* Molecular mechanisms governing complement activation in neuromyelitis optica. *J. Neuroimmunol.* **275**, 13–14 (2014).

74. Kalogeris, T., Baines, C. P., Krenz, M. & Korthuis, R. J. Cell Biology of Ischemia/Reperfusion Injury. in *International Review of Cell and Molecular Biology* 271–284 (Elsevier Inc., 2012). doi:10.1016/B978-0-12-394309-5.00006-7

75. Sárvári, M. *et al.* Inhibition of C1q-h-amyloid binding protects hippocampal cells against complement mediated toxicity. *J. Neuroimmunol.* **137**, 12–18 (2003).

76. Morgan, B. P. & Harris, C. L. Complement, a target for therapy in inflammatory and degenerative diseases. *Nature Reviews Drug Discovery* **14**, 857–877 (2015).

77. Pardridge, W. M. The Blood-Brain Barrier: Bottleneck in Brain Drug Development. *J.*

Am. Soc. Exp. Neurother. **2**, 3–14 (2005).

78. Veerhuis, R., Nielsen, H. M. & Tenner, A. J. Complement in the brain. *Molecular Immunology* (2011). doi:10.1016/j.molimm.2011.04.003

79. Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews* **64**, 4–17 (2012).

80. Congreve, M., Carr, R., Murray, C. & Jhoti, H. A 'Rule of Three' for fragment-based lead discovery? *Drug Discovery Today* (2003). doi:10.1016/S1359-6446(03)02831-9

81. How Does Surface Plasmon Resonance Work? | Biosensing Instrument. Available at: https://biosensingusa.com/technologies/surface-plasmon-resonance/surface-plasmon-resonance-work/. (Accessed: 22nd November 2019)

82. Principle and Protocol of Surface Plasmon Resonance (SPR) - Creative BioMart. Available at: https://www.creativebiomart.net/resource/principle-protocol-principle-and-protocol-of-surface-plasmon-resonance-spr-361.htm. (Accessed: 22nd November 2019)

83. Smyth, M. S. & Martin, J. H. J. *Review x Ray crystallography*. *J Clin Pathol: Mol Pathol* **53**, (2000).

84. Garcia, B. L., Zwarthoff, S. A., Rooijakkers, S. H. M. & Geisbrecht, B. V. Novel Evasion Mechanisms of the Classical Complement Pathway. *J. Immunol.* **197**, 2051–2060 (2016).

85. Xie, J. *et al.* Structural determination of the complement inhibitory domain of Borrelia burgdorferi BBK32 provides insight into classical pathway complement evasion by Lyme disease spirochetes. *PLOS Pathog.* 1–29 (2019). doi:10.1371/journal.ppat.1007659

APPENDIX

Surface plasmon resonance

Surface plasmon resonance (SPR) is an instrument used to quantify and analyze molecular binding interactions between an immobilized protein and a solubilized, mobile analyte. This is achieved by measuring the change in reflected intensity of a light beam upon interaction with the metal film upon which the ligand is immobilized. SPR utilizes the differential refractive indices of a prism, through which polarized light is projected, and the metal film (typically gold or silver) on the surface of the sensor chip [81].

The protein of interest is immobilized onto a sensor chip by amine-coupling reaction. The CM5 sensor chip is inserted into the Biacore T200 and N-Hydroxysuccinimide (NHS)/ N-ethyl-N'-(dimethylaminopropyl) (EDC) is injected over the surface to activate the chip by modification of carboxymethyl groups to NHS esters [82]. The activation of the chip surface is recorded as a baseline and the ligand is then injected over the chip. The N-terminus and ε-amino groups of lysine residues on the ligand covalently bind the sensor surface as the NHS esters of the chip react with the amino groups of the ligand. Unbound/electrostatically bound ligand and unreacted NHS-esters are eluted and deactivated by ethanolamine pH 8.5 and 10 mM NaOH [81].

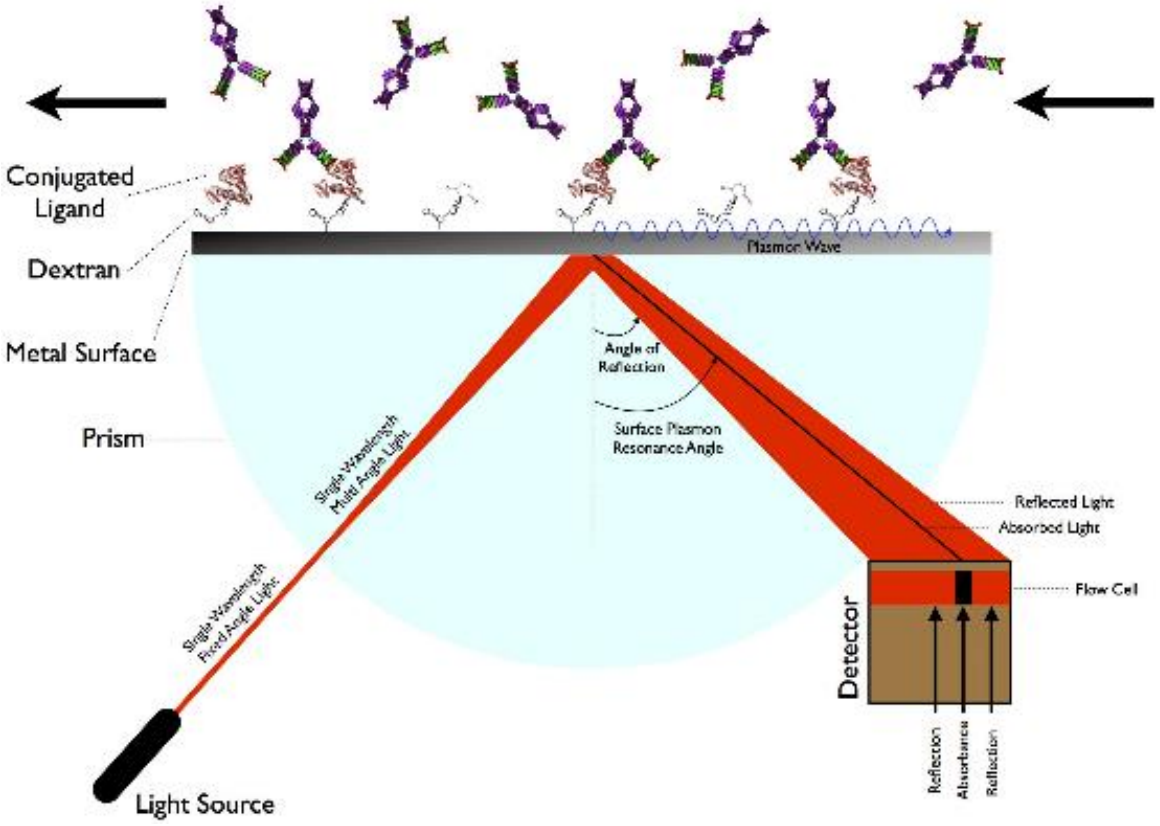
Varying concentrations of the analyte, as well as a blank, are then injected over the chip surface with immobilized ligand over multiple cycles. Polarized light travels through the glass prism and is focused onto a particular point on the metal film of the sensor surface at a set resonance angle [82]. At the critical angle, all out-going reflected light is reflected internally (total internal reflection). With total internal reflection, the light is absorbed and excites the electrons that exist at the interface between the metal surface and prism and converts them to surface

plasmons, which is the particle name of the electron density wave [82]. This conversion creates a dip or gap in the intensity of the reflected light. During a binding event between the analyte and ligand, there is an angular shift in position from the original dip and the dip that occurs during molecular binding due to the conformational change in the reactants bound to the chip surface [81].

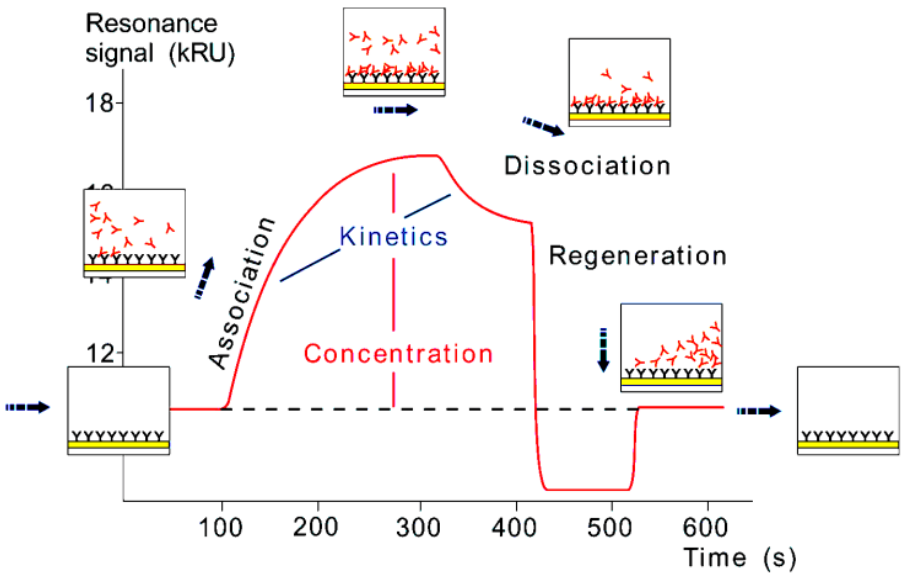
The binding of reactants on the chip surface also alters the refractive index, which is measured by the change in the resonance angle. The change in refractive index is proportional to the concentration of immobilized ligand, converted into arbitrary resonance units (RU) [81]. During a binding event, the RU increases, and after a predetermined period of time, buffer is injected over the chip to dissociate the analyte-ligand complex, thereby dropping the RU. This angular shift versus time allows for quantification of binding kinetics between the target ligand and analyte over varying concentrations, yielding dose-dependent binding kinetics [81]. The maximal response (R_{max}) is calculated for the immobilized ligand by assuming a binding event of 1:1 with the analyte [82].

The interaction between the analyte and ligand is graphically represented by the association event, steady state, and the dissociation event over time with no changes to the molecular properties of either reactant [81]. With a high dissociation rate ($<10^{-1}$), steady state experiments are used to calculate the equilibrium dissociation constant by dividing the dissociation rate by the association rate. Multi-cycle kinetics can also be used to define the interactions between reactants by injecting the analyte over the chip surface in separate cycles and their resultant curves overlaid into one sensorgram [81].

A


Conjugated Ligand
Dextran
Metal Surface
Prism
Light Source
Plasmon Wave
Angle of Reflection
Surface Plasmon Resonance Angle
Reflected Light
Absorbed Light
Flow Cell
Detector
Reflection
Absorbance
Reflection

B


Resonance signal (kRU)
18
12
Association
Kinetics
Concentration
Dissociation
Regeneration
100
200
300
400
500
600
Time (s)

Appendix Figure 1. SPR schematic of instrument mechanisms and enzyme kinetic analysis. (A) Molecular binding interactions between protein and analyte is measured by the change in the reflected intensity of a light beam upon interaction with the metal film upon which the ligand is immobilized. Utilizing the differential refractive indices of a prism, through which polarized light is projected, and the metal film (typically gold or silver) on the surface of the sensor chip. Polarized light is focused onto the chip surface, and during a molecule binding event, a change in refractive index is marked by an angular shift in the intensity of reflected light [80]. **(B)** The interaction between the analyte and ligand is graphically represented by the association event, steady state, dissociation event and subsequent regeneration over time, quantified by resonance units (RU) [79].

Complement inhibition assays

Complement inhibition in this study was measured using a series of assays that looked at classical complement pathway inhibition via downstream complement product deposition, direct inhibition of the initiating proteases, C1r or C1s on a peptide substrate, or complement-mediated hemolysis of sheep red blood cells. Classical pathway inhibition measured by C4b deposition utilizes an enzyme-linked immunosorbent assay (ELISA) to assess the presence of proteins by antibodies. Because complement is normally found circulating in normal human serum at relatively high concentrations, complement fixation products must be bound and immobilized to quantify deposition and, ultimately, complement activation.

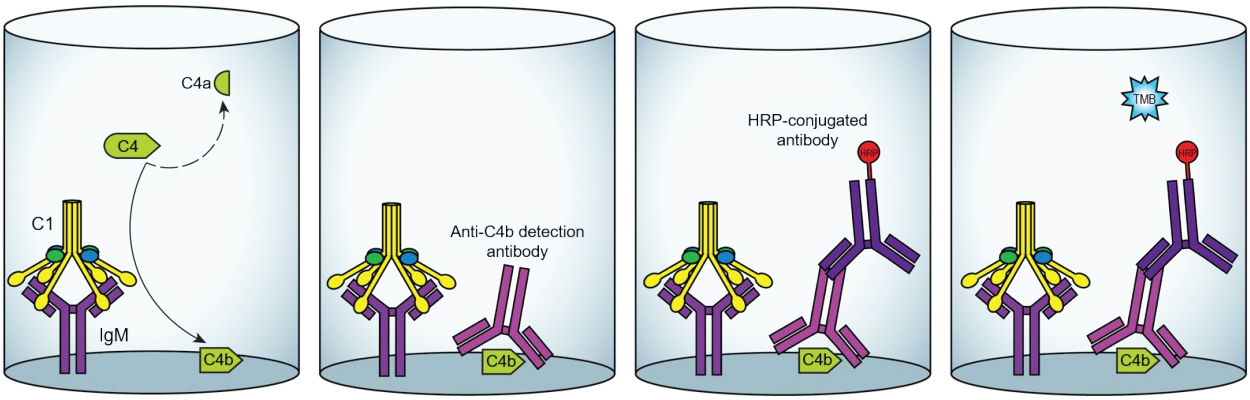
To facilitate complement fixation, IgM antibody is immobilized on a high-binding, 96-well plate. The remaining open surfaces of the wells are then blocked with 1% bovine serum

albumin to prevent non-specific binding of complement proteins or antibody. When 2% normal human serum is added to the microtiter wells, the pentameric structure of IgM is targeted by recognition molecule, C1q, within the C1 complex. A potent and specific inhibitor of C1 proteases simultaneously dispensed into solution would block the subsequent activation of bound C1, thereby halting the proteolytic cascade from forward progression. However, if C1 activation is not impeded, C1r autoactivates and cleaves zymogen C1s. C1s cleaves C4 into a trimer and C2 into its dimer subunits. While C4a, C4c, and C2a are released into fluid-phase, C4b and C2b remain bound together at the plate surface as C3-convertase.

While downstream complement deposition products can also be measured, isolation of C4b by mouse anti-C4b hybrid antibody eliminates alternative pathway from consideration, so that only classical complement activation (and/or lectin pathway) is measured. With C4b and murine anti-C4 hybrid antibody in complex at the plate surface, secondary anti-mouse goat antibody conjugated to horseradish peroxidase (HRP) is added to bind the existing immune complex and amplify signal detection of C4b by 3,3′,5,5′-Tetramethylbenzidine (TMB), the colorimetric reagent used for visualization of target proteins. To prevent undesirable continued reaction progression and increased background signal, sulfuric acid is added as the stop reagent, changing to solution from blue to yellow.

Differential optical density read by spectrophotometer microplate reader at 450 nm represents varying degrees of complement activation and thus complement inactivation by compound inhibitor. By serially diluting the inhibitor concentration, dose-dependent inhibition can be assessed. Serving as controls for the assay are two rows dedicated to either 100% complement activation, in which there is no inhibitor, or 0% complement activation, in which

there is neither normal human serum nor inhibitor but only classical pathway-specific assay buffer.



Appendix Figure 2. ELISA-based assay measuring complement activity through component fixation. IgM is immobilized onto a high-binding plate surface and blocked from non-specific binding. Normal human serum containing complement is introduced to the plate with the presence of an inhibitor. If complement is active, C1q will bind IgM and activate the CP. Antibody against complement fixation products is introduced and a secondary HRP-conjugated antibody is added. With the presentation of TMB into solution, the presence and concentration of complement deposition fragments is marked by a colorimetric change that can be visualized and quantified by spectrophotometer.

The second method by which to quantify C1r or C1s inactivation is by inhibition of enzyme-mediated cleavage of a peptide substrate specific for that protease by the inhibitory compound under investigation. In our study, we found that peptide, Z-Gly-Arg-SBzl, provided the optimal substrate for C1 protease to act upon.

In a 96-well conical plate, 500 µM final concentration of enzyme was added concurrently with 50 mM Tris buffer pH 8.0, and 500 µM inhibitor to allow for inhibitor-enzyme complex

formation. Separately, and in their own reservoir, peptide and Ellman's reagent (5,5'-dithiobis-(2-nitrobenzoic acid) or DTNB) are mixed. DTNB is a stoichiometric measure yielding colorimetric changes during cleavage events measured by OD detection by spectrophotometer at 412 nm. The peptide-DTNB mixture is then added to the enzyme-inhibitor solution, providing rapid color change from the reaction of DTNB to substrate cleavage. Greater absorbance values equate to lesser C1 protease inhibition by the compound. To determine maximal signal, one row is designated as the positive control without any inhibitor present in solution. The negative control row contains no enzyme to account for background noise.

X-Ray Crystallography

X-ray crystallography utilizes the diffraction patterns of a crystal sample to yield a three-dimensional molecular structure for making structural determinations of proteins or biological macromolecules. Protein crystallography relies on the high quality and high concentration of homogenous, soluble protein. Crystals are grown from protein solutions containing a minimum protein concentration of 2 mg/mL that are induced to slowly precipitate out of solution in the form of a crystal under given crystallization conditions [83]. These conditions can be obtained through commercially available "crystal screen" packages containing 50 solutions of differing precipitant, buffer, pH, and salt [83]. The leading conditions in our study for the crystallization of wild-type C1r-2SP were 0.2 M sodium acetate trihydrate, 0.1 M Tris hydrochloride pH 8.5, and 30% w/v polyethylene glycol 4,000 in one and the second containing 0.2 M magnesium chloride hexahydrate, 0.1 M Tris hydrochloride pH 8.5, 30% w/v polyethylene glycol 4,000.

Crystallization of the protein occurs through vapor diffusion of the 2 µL protein drop solution exchanged with the 50 µL crystallization solution. There are two techniques for establishing vapor diffusion. One method is manually introducing a "hanging drop" of the

sample suspended from a glass coverslip over a reservoir of the crystallization condition while the other method is mechanically establishing vapor diffusion through a "sitting drop," which utilizes a Crystal Gryphon instrument that dispenses the proteins into 96-channel crystal plates containing the crystallization condition [83]. To co-crystallize a compound in complex with the protein, small volumes of compound solution are added to the protein sample prior to vapor diffusion.

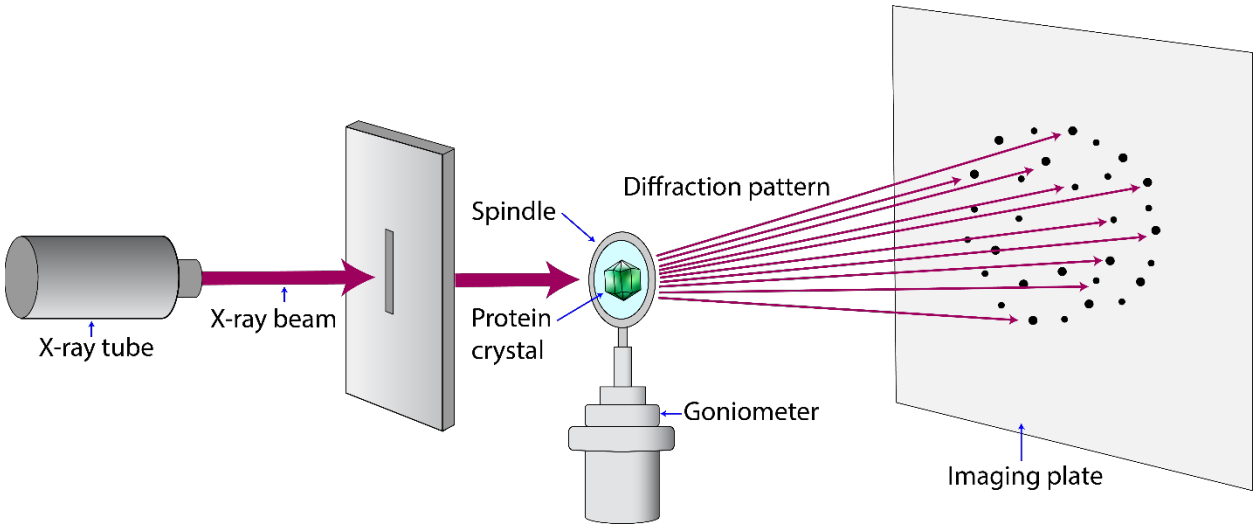
Crystal growth can take days or weeks at room temperature, but once crystals are obtained, they are introduced into a cryoprotectant buffer so that they may be collected and stored in liquid nitrogen to prevent degradation. In our study, our proteins are then sent to the Southeast Regional Collaborative Access Team (SER-CAT) located at the Advanced Photon Source at the Argonne National Laboratory for diffraction analysis.

When the crystals are ready for analysis of their x-ray diffraction behavior, they are remotely and mechanically mounted within a rotating spindle and adjusted for proper alignment with the x-ray beam using what is called a goniometer head. The x-rays are generated by accelerating electrons in a synchrotron storage ring and the x-rays focused into a beam. The diameter of the incident beam, as well as alignment of the crystal must be carefully adjusted to ensure proper exposure and prevent radiation damage [83].

The crystal rotates within the spindle perpendicularly to the incident beam to provide the maximum amount of diffraction data. After exposure, diffraction spots are collected on an imaging plate revealing a digitized image. The diffraction spots are resolved based on the distance from the crystal to the detector with the highest resolution accounting for the edge of the detection. The diffraction image becomes strongest at lower resolutions, with a resolution of 3 Å sufficient for detecting amino acid side chains in an electron density map. The resolution of the

image is dictated by how well-ordered the molecular subunit is. Within the crystal, the unit cell is defined as the smallest repeating unit that makes up the crystal with dimensions given by three lengths and angles. The dimensions of the unit cells are proportional to the number of diffraction spots for a given unit area. The space group of the crystal is given by the symmetry of the diffraction pattern, which is defined by the configuration and packing of the molecules within the crystal lattice. The more symmetrical a crystal structure, the more ordered the diffraction spots appear in the image [83].

Computer programs with well-established algorithms for making structural determinations are employed to resolve the unit cell dimensions and orientation, measure the intensity of diffraction spots, and normalize spot intensity using a scale factor to relate all the information within the data set. Given the availability of published models of related protein structures, molecular replacement is utilized to reconstruct amino acid sequences within the protein. Combining the "borrowed" model and diffraction data produces electron density maps that account for the location and bond angles of amino acid sequences of the protein. The crystal structure must undergo several rounds of model building and refinement to establish the structural molecular nuances of a given protein. The quality of the protein model is calculated from the reflection data and given by an R-factor [83]. The lower the R-value, the closer the fit is to "perfect." In general, a random set of atoms produces an R-value of approximately 0.63, whereas high-quality published structures aim for values of about 0.20.



Appendix Figure 3. Protein x-ray crystallography schematic. An x-ray beam generated from the synchrotron storage ring is directed at the protein crystal that has been properly aligned within its rotating spindle using the goniometer head. The resultant diffraction pattern is visualized on an imaging plate to yield protein structural determinations.